

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 21:08:50 ; Search time 5576 Seconds
(without alignments)
10827.985 Million cell updates/sec

Title: US-10-099-704-1
Perfect score: 1393
Sequence: 1 ccgaagtgtgcaatggag.....cgaaaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb.ba.*
 - 2: gb.htg.*
 - 3: gb.in.*
 - 4: gb.om.*
 - 5: gb.ov.*
 - 6: gb.pat.*
 - 7: gb.ph.*
 - 8: gb.pl.*
 - 9: gb.pr.*
 - 10: gb.ro.*
 - 11: gb.sts.*
 - 12: gb.sy.*
 - 13: gb.un.*
 - 14: gb.vi.*
 - 15: em.ba.*
 - 16: em.fun.*
 - 17: em.hum.*
 - 18: em.in.*
 - 19: em.mu.*
 - 20: em.om.*
 - 21: em.or.*
 - 22: em.ov.*
 - 23: em.pat.*
 - 24: em.ph.*
 - 25: em.pl.*
 - 26: em.ro.*
 - 27: em.sts.*
 - 28: em.un.*
 - 29: em.vi.*
 - 30: em.htg.hum.*
 - 31: em.htg.inv.*
 - 32: em.htg.other.*
 - 33: em.htg.mus.*
 - 34: em.htg.pln.*
 - 35: em.htg.rod.*
 - 36: em.htg.man.*
 - 37: em.htg.vrt.*
 - 38: em.sy.*
 - 39: em.htgo.hum.*
 - 40: em.htgo.mus.*
 - 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1393	100.0	1393	6	BD236906	BD236906 Method fo
2	1393	100.0	1393	6	AR208625	AR208625 Sequence
3	214.8	15.4	2505	8	CPU312754	AJ312754 Claviceps
4	210.8	15.1	2182	8	CPU011963	AJ011963 Claviceps
5	210.8	15.1	2299	8	AY259840	AY259840 Claviceps
6	210.6	15.1	2694	8	CPU312753	AJ312753 Claviceps
7	205.2	14.7	1347	6	AR181512	AR181512 Sequence
8	205.2	14.7	1353	6	AR181513	AR181513 Sequence
9	194	13.9	2143	8	AY262013	AY262013 Balansa
10	194	13.9	2570	8	CPU312752	AJ312752 Claviceps
11	193.6	13.9	1783	8	AY262014	AY262014 Claviceps
12	182.6	13.1	1894	8	CLCDMAW	IJ3640 Claviceps f
13	176.6	12.7	1598	6	AR181515	AR181515 Sequence
14	176.6	12.7	6055	8	AY259837	AY259837 Neotyphod
15	176.6	12.7	6640	8	AY259837	AY259837 Epichloe
16	173.4	12.4	1908	6	AR181514	AR181514 Sequence
17	173.4	12.4	6702	8	AY259838	AY259838 Neotyphod
C 18	44.6	3.2	125020	9	AF429315	AF429315 Homo sapi
C 19	43.6	3.1	2000	6	AX655393	AX655393 Sequence
C 20	41.6	3.0	125020	9	AF429315	AF429315 Homo sapi
C 21	41	2.9	115368	9	AC135049	AC135049 Homo sapi
C 22	41	2.9	148054	9	AC026467	AC026467 Homo sapi
C 23	41	2.9	156899	2	AC103884	AC103884 Homo sapi
C 24	41	2.9	173980	9	AC116552	AC116552 Homo sapi
C 25	41	2.9	190590	2	AC135045	AC135045 Homo sapi
C 26	40.2	2.9	251271	2	AC129644	AC129644 Rattus no
C 27	40.2	2.9	253942	2	AC096037	AC096037 Rattus no
C 28	40.2	2.9	273754	2	AC120615	AC120615 Rattus no
C 29	38.8	2.8	2000	6	AX655393	AX655393 Sequence
C 30	38	2.7	295500	1	AP005954	AP005954 Bradyrhiz
C 31	37.8	2.7	2830	3	AF027972	AF027972 Nephila c
C 32	37.8	2.7	2830	6	AR091362	AR091362 Sequence
C 33	37.8	2.7	111738	2	AL592304	AL592304 Homo sapi
C 34	37.8	2.7	136116	2	AC110874	AC110874 Gallus ga
C 35	37.8	2.7	140207	9	AL356390	AL356390 Human DNA
C 36	37.8	2.7	166471	2	AC091052	AC091052 Gallus ga
C 37	37.8	2.7	202272	2	AC146292	AC146292 Homo sapi
C 38	37.8	2.7	239116	9	AP000426	AP000426 Homo sapi
C 39	37.6	2.7	112533	9	AL139097	AL139097 Human DNA
C 40	37.6	2.7	177521	2	AL161438	AL161438 Homo sapi
C 41	37.2	2.7	2765	8	AK064973	AK064973 Oryza sat
C 42	37.2	2.7	2997	8	AK122184	AK122184 Oryza sat
C 43	37.2	2.7	223951	2	EX005010	EX005010 Danio rer
C 44	36.6	2.6	1594	3	AY132363	AY132363 Spiroanucl
C 45	36.6	2.6	176099	9	AC108201	AC108201 Homo sapi

ALIGNMENTS

RESULT 1
BD236906
LOCUS BD236906 1393 bp DNA linear PAT 17-JUL-2003
DEFINITION Method for producing polypeptides in Aspergillus variant cells.
ACCESSION BD236906
VERSION BD236906.1 GI:33046676
KEYWORDS JP 2002533133-A/1.
SOURCE Aspergillus oryzae
ORGANISM Aspergillus oryzae
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
REFERENCE 1 (bases 1 to 1393)
AUTHORS Christensen,B.E., Moellgaard,H., Kaasgaard,S. and Lehmbeck,J.
TITLE Method for producing polypeptides in Aspergillus variant cells
JOURNAL Patent: JP 2002533133-A 1 08-OCT-2002;

NOVOZYMES AS
OS Aspergillus oryzae
PN JP 2002533133-A/1
PD 08-OCT-2002
PR 22-DEC-1999 JP 2000591212
PR 23-DEC-1998 DK PA 199801726, 27-MAY-1999 DK PA 199900745 PI
BOERN EGGERT CHRISTENSEN, HENRIK MOELIGAARD, SVEND KAASGAARD, PI
JAN LEHMBECK
PC
C12N15/09, C12N1/15, C12N9/00, C12N9/20, C12N9/42, C12N15/01, C12P21/ PC
02//
PC (C12N1/15, C12P1/69), C12N15/00, C12N15/00
CC Method for producing polypeptides in Aspergillus variant cells
FH Key Location/Qualifiers
FT CDS (15)..(1328).
FEATURES
source
1. .1393
/organism="Aspergillus oryzae"
/mol_type="genomic DNA"
/db_xref="taxon:5062"
ORIGIN
Query Match 100.0%; Score 1393; DB 6; Length 1393;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGAAGCTGACGAATGGAGATCTCCAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT 60
DB 1 CCGAAGCTGACGAATGGAGATCTCCAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT 60
QY 61 ACCTGCTGAGTCAAGCCCTGAACCTCTCGAAACAGGACCAACAAATGGTGGTATAGCA 120
DB 61 ACCTGCTGAGTCAAGCCCTGAACCTCTCGAAACAGGACCAACAAATGGTGGTATAGCA 120
QY 121 CAGCTCGGATGTTGCCACCATGATGGCGGGCGCGGTATGATGTTTCACGACAGTACA 180
DB 121 CAGCTCGGATGTTGCCACCATGATGGCGGGCGCGGTATGATGTTTCACGACAGTACA 180
QY 181 AGTTCTCTGATCCACCGTGAGTCAATCATCCCGCGGTGGTCCATACCCAGAAAAGG 240
DB 181 AGTTCTCTGATCCACCGTGAGTCAATCATCCCGCGGTGGTCCATACCCAGAAAAGG 240
QY 241 GTGAGCCCATGCACTGGAAGAGTCACTCACACGCTTCGGAATTCCTTTTCGAGCTGAGCT 300
DB 241 GTGAGCCCATGCACTGGAAGAGTCACTCACACGCTTCGGAATTCCTTTTCGAGCTGAGCT 300
QY 301 TCAATTACTCCAAATCACTACTACGTTTGCATTCGAGCCCTCCGTTCCCTGACGGGAA 360
DB 301 TCAATTACTCCAAATCACTACTACGTTTGCATTCGAGCCCTCCGTTCCCTGACGGGAA 360
QY 361 CGAAGGATGATCCATTCACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGGCA 420
DB 361 CGAAGGATGATCCATTCACACCCAGGCAATCAGGCTGTTCTCCAGGACCTCAAGGCA 420
QY 421 TGGTTCAGGGCTTGACCTGGATGTTGATCAATTCATTAAGCATTTGGTTCGTTTCGG 480
DB 421 TGGTTCAGGGCTTGACCTGGATGTTGATCAATTCATTAAGCATTTGGTTCGTTTCGG 480
QY 481 AGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGTCTTCAAGACACAGA 540
DB 481 AGGAAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGTCTTCAAGACACAGA 540
QY 541 ACAAACTGGACCGGATCTGGAGCCATCTGGGATATTGTTTGAAGACCTACATCTACC 600
DB 541 ACAAACTGGACCGGATCTGGAGCCATCTGGGATATTGTTTGAAGACCTACATCTACC 600
QY 601 CGGGGATCAAGTGCATCGGACCGGACCCCAAGAGAGAGTCTATGTTTGAAGCAATCA 660
DB 601 CGGGGATCAAGTGCATCGGACCGGACCCCAAGAGAGAGTCTATGTTTGAAGCAATCA 660
QY 661 AGGCTCCGACAAAGTTGGCAAGTTGCGCAATCTCCACTGGCAATCTCCGAGGAGTTTATAG 720
DB 661 AGGCTCCGACAAAGTTGGCAAGTTGCGCAATCTCCACTGGCAATCTCCGAGGAGTTTATAG 720

QY 721 CTGAGCGAGCAGCCACCCCTCTCGGCCACTTTCTCTCATCGATTTGGTCAAGCCGTCG 780
DB 721 CTGAGCGAGCAGCCACCCCTCTCGGCCACTTTCTCTCATCGATTTGGTCAAGCCGTCG 780
QY 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTGGCCCTCCATCGAAGTA 840
DB 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTGGCCCTCCATCGAAGTA 840
QY 841 TTTGGACTCTCAACGGCGAGCGGAACCATCCAGAGACACTGGATGGTCTCGATGCGGTGA 900
DB 841 TTTGGACTCTCAACGGCGAGCGGAACCATCCAGAGACACTGGATGGTCTCGATGCGGTGA 900
QY 901 GGGAGCTGTGGCAGCTATTGCCCGTACGAGGAGGTCTGTCCACTGCCGAACTGCTTT 960
DB 901 GGGAGCTGTGGCAGCTATTGCCCGTACGAGGAGGTCTGTCCACTGCCGAACTGCTTT 960
QY 961 ACAGCGCGGTACCTCACCGCAGGAGCAGTCCCTTCAATATAAATTTTACCTTGTCTC 1020
DB 961 ACAGCGCGGTACCTCACCGCAGGAGCAGTCCCTTCAATATAAATTTTACCTTGTCTC 1020
QY 1021 CTAAGAGCGCACTTCCCGAACCAAGATCTATTTCCCTGCTTTTGGGCGAAGCAAAAA 1080
DB 1021 CTAAGAGCGCACTTCCCGAACCAAGATCTATTTCCCTGCTTTTGGGCGAAGCAAAAA 1080
QY 1081 CCATCGCGGAGGATTGGCCACCTCTTTGAGAGCAGAGGTTCGGGTGGCTTGGCTAAGA 1140
DB 1081 CCATCGCGGAGGATTGGCCACCTCTTTGAGAGCAGAGGTTCGGGTGGCTTGGCTAAGA 1140
QY 1141 GCTATCAGCGGATTTGGCATCTCTATCTATCCGATGCGACCTCGACACGCAATCACC 1200
DB 1141 GCTATCAGCGGATTTGGCATCTCTATCTATCCGATGCGACCTCGACACGCAATCACC 1200
QY 1201 TGGAGCGGTGGATCTCTCTTCTACAGGGGAAAAACCGTACATGATGATGCTACTCC 1260
DB 1201 TGGAGCGGTGGATCTCTCTTCTTACAGGGGAAAAACCGTACATGATGATGCTACTCC 1260
QY 1261 ATACCTTCGAAGCGTTCAGTCTGCTGCCAGGAGGTGGCTATGTCACGATGGCCACA 1320
DB 1261 ATACCTTCGAAGCGTTCAGTCTGCTGCCAGGAGGTGGCTATGTCACGATGGCCACA 1320
QY 1321 ATCCTTAGGACTAGTTTATCCCTTCAATCTATGATCCCTGGTGAATGTTGGTCGAAAAA 1380
DB 1321 ATCCTTAGGACTAGTTTATCCCTTCAATCTATGATCCCTGGTGAATGTTGGTCGAAAAA 1380
QY 1381 AAAAAAAAAAAAAA 1393
DB 1381 AAAAAAAAAAAAAA 1393
RESULT 2
LOCUS AR208625 1393 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383781.
ACCESSION AR208625
VERSION AR208625.1 GI:21509828
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1393)
AUTHORS Christensen, B. Eggert., Mollgaard, H., Kaasgaard, S. and Lehmbek, J.
TITLE Methods for producing polypeptides in aspergillus mutant cells
JOURNAL Patent: US 6383781-A 1 07-MAY-2002;
FEATURES
Location/Qualifiers
source
1. .1393
/organism="unknown"
/mol_type="unassigned DNA"

QY 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT 60
DB 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAACACTGCTGCCAAAGCCCTTCT 60
QY 61 ACCTGCTGAGTCAAGCCCTGAACCTCTCGAAACAGAGCAGCAACAAATGGTGTATAGCA 120
DB 61 ACCTGCTGAGTCAAGCCCTGAACCTCTCGAAACAGAGCAGCAACAAATGGTGTATAGCA 120
QY 121 CAGCTCCGATGTTTGGCCACATGATCGCGGGGCGGCTATGATGTTACAGCAGATACA 180
DB 121 CAGCTCCGATGTTTGGCCACATGATGCGGGGCGGCTATGATGTTACAGCAGATACA 180
QY 181 AGTTCTCTGTATPCCACCGTGAAGTCAATCATCCCGCGTGGGTCCATACCCAGAAAAGG 240
DB 181 AGTTCTCTGTATPCCACCGTGAAGTCAATCATCCCGCGTGGGTCCATACCCAGAAAAGG 240
QY 241 GTCAGGCCATGCACTCGAGAGATCATCTCAGCGTTCCGACTTCCTTTTCGAGCTGAGCT 300
DB 241 GTCAGGCCATGCACTCGAGAGATCATCTCAGCGTTCCGACTTCCTTTTCGAGCTGAGCT 300
QY 301 TCAATTACTCCAAATCACTACTACGGTTCGATTCGAGCCCTCGGTTCCCTGACGGGAA 360
DB 301 TCAATTACTCCAAATCACTACTACGGTTCGATTCGAGCCCTCGGTTCCCTGACGGGAA 360
QY 361 CGAAGGATGATCCATTCACACCCAGGCAATCAGGCTGTTCTCAGGACCTCAAGGCCA 420
DB 361 CGAAGGATGATCCATTCACACCCAGGCAATCAGGCTGTTCTCAGGACCTCAAGGCCA 420
QY 421 TGGTTCAGGGCTTGACCTGGAATGGTTCGATCATTTCACTAAAGCATTTGGTTCGTTCCG 480
DB 421 TGGTTCAGGGCTTGACCTGGAATGGTTCGATCATTTCACTAAAGCATTTGGTTCGTTCCG 480
QY 481 AGGAAGGCTCGGATCTCTGATGATCGAGATATTGAGATCCCGCTTCACAGACACAGA 540
DB 481 AGGAAGGCTCGGATCTCTGATGATCGAGATATTGAGATCCCGCTTCACAGACACAGA 540
QY 541 ACAAATGGCAGCGGATCTGAGCCATCTGGCGATATTGTTGAAGACTCATCTACC 600
DB 541 ACAAATGGCAGCGGATCTGAGCCATCTGGCGATATTGTTGAAGACTCATCTACC 600
QY 601 CGCGGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 660
DB 601 CGCGGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 660
QY 661 AGGCTCGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCA 720
DB 661 AGGCTCGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCA 720
QY 721 CTGAGCGAGCACCCCTCTCGGCACTTTCTCTCATGCGATTTGGTCAAGCCCTCGG 780
DB 721 CTGAGCGAGCACCCCTCTCGGCACTTTCTCTCATGCGATTTGGTCAAGCCCTCGG 780
QY 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTGGCTCCATCGAAGGTA 840
DB 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGAGCTCGACCTGGCTCCATCGAAGGTA 840
QY 841 TTTTGGCTCTCAAGCGGCGAGGAAAGATCCAGAGACACTGGATGTTGGATGGCGCTCA 900
DB 841 TTTTGGCTCTCAAGCGGCGAGGAAAGATCCAGAGACACTGGATGTTGGATGGCGCTCA 900
QY 901 GGGAGCTGTGGAGCTATTCCCGCTCAGGAGGCTGTGTCCACTGCCAACTGCTTTT 960
DB 901 GGGAGCTGTGGAGCTATTCCCGCTCAGGAGGCTGTGTCCACTGCCAACTGCTTTT 960
QY 961 ACAGCGGGTACTCAGCGGAGGAGGCTCCCTTCATTATAAATTTTACCTGCTC 1020
DB 961 ACAGCGGGTACTCAGCGGAGGAGGCTCCCTTCATTATAAATTTTACCTGCTC 1020
QY 1021 CTAAGACGCACTTCCGAAACCAAGATCTATTTCCTGCTTTTGGGCGAGAACGACAAA 1080
DB 1021 CTAAGACGCACTTCCGAAACCAAGATCTATTTCCTGCTTTTGGGCGAGAACGACAAA 1080

QY 1081 CCATCGCGGAGGATTCGCCACCTTCTTTGAGAGCAGAGTGGGTGGCTTGAAGA 1140
DB 1081 CCATCGCGGAGGATTCGCCACCTTCTTTGAGAGCAGAGTGGGTGGCTTGAAGA 1140
QY 1141 GCTATCCAGCGGATTCGCATCCCTACTATCCGATGTGGACCTGCAGACCGCAATCACC 1200
DB 1141 GCTATCCAGCGGATTCGCATCCCTACTATCCGATGTGGACCTGCAGACCGCAATCACC 1200
QY 1201 TGCAGGCGTGATCTCCTTCTTTTACAAGGGGAAAAACCGTACATGAGTGTGTACCTCC 1260
DB 1201 TGCAGGCGTGATCTCCTTCTTTTACAAGGGGAAAAACCGTACATGAGTGTGTACCTCC 1260
QY 1261 ATACCTCGAAGCGTTCACTGCTCTGCCAGGAGTGGCTATGTGTACGATGGCCACA 1320
DB 1261 ATACCTCGAAGCGTTCACTGCTCTGCCAGGAGTGGCTATGTGTACGATGGCCACA 1320
QY 1321 ATCCTTAGGACTAGTTCCTTTCATTTATGATCATCCGTTGATGTGTGGTGGAAAAA 1380
DB 1321 ATCCTTAGGACTAGTTCCTTTCATTTATGATCATCCGTTGATGTGTGGTGGAAAAA 1380
QY 1381 AAAAAA 1393
DB 1381 AAAAAA 1393

RESULT 3
CPU312754 2505 bp DNA linear PLN 24-APR-2001
LOCUS Claviceps purpurea partial cpd1 gene for putative
DEFINITION dimethyl-allyl-tryptophan-synthase, exons 1-3, strain T5.
ACCESSION AJ312754
VERSION AJ312754.1 GI:13810199
KEYWORDS cpd1 gene; dimethyl-allyl-tryptophan-synthase.
SOURCE Claviceps purpurea (ergot fungus)
ORGANISM Claviceps purpurea
REFERENCE 1 Correira, T. and Tudzynski, P.
AUTHORS Molecular analysis of dimethyl-allyl-tryptophan-synthase-genes
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 2505)
AUTHORS Tudzynski, P.
TUDZYNSKI, P.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2001) Tudzynski P., Institut fuer Botanik,
Westfaelische Wilhelms-Universitaet, Schlossgarten 3, 48149
Muenster, GERMANY
FEATURES
source
1. .2505
/organism="Claviceps purpurea"
/mol_type="genomic DNA"
/strain="T5"
/db_xref="taxon:5111"
/country="Germany"
1064. .2505
/gene="cpd1"
join(1064. .2229, 2295. .2351, 2406. .2505)
/gene="cpd1"
/function="isoprenylation of tryptophane"
/codon_start=1
/product="putative dimethyl-allyl-tryptophan-synthase"
/protein_id="CAC37397.1"
/db_xref="GI:13810200"
/db_xref="SPTREMBL:Q9C140"
/translation="MSTAKDPGVSYLSLIFDPFSGNEQRLMWHSTAPFAALMDNA
GYNTHDOYRHGIFKKHIIPLGVVPTKDKERWLSILTRCGLPLELSINCDTSVVRVT
YEPINEVGTGKDPNTLIMASVOKLAQIQAGIDLEWFSYFKDELTLDESATLQS
NLEKQVKTQKIALDLKESQFALKYVFPPLASITATGKSTHLLI FDSFKLSQKHD
SIQXAXQVLCDYVSRNHSAESDQIHAKLLSCDLIDPAKSRVLIYLLKTVLSLV
MEDLWTGGQRVDASTMDGLMRELWLLKVPHTGHLEYPKGYLELGEIIPNEQLPSVA
NYTLHNHNPPEQVYFTFGMDAEISNLTIFQRHFDMAKRYVFLQDSYPVH
DFESINLYHAYISFSYMPVSGSTELYQHGLRKLNNFLVADSPISFDAYR"
1064. .2229

gene
CDS
exon

```

/ gene="cpd1"
/ number=1
2230.2294
/ gene="cpd1"
/ number=1
2295.2351
/ gene="cpd1"
/ number=2
2352.2405
/ gene="cpd1"
/ number=2
2406.2505
/ gene="cpd1"
/ number=3

ORIGIN
Query Match          15.4%; Score 214.8; DB 8; Length 2505;
Best Local Similarity 52.3%; Pred. No. 1.7e-52;
Matches 586; Conservative 2; Mismatches 499; Indels 33; Gaps 4;

Qy 57 TTCTAGTGTGAGTCAAGCCCTGAACCTCTCGAAGACGACACACAAAATGGTGGTAT 116
Db 1097 TAGGAAATTTTGATGTCTGATTTTGGATTTCCCGAGCAATGAACAGCGACTATGGTGGCAC 1156

Qy 117 AGCAGACTCCGATGTTTGGCCACCATGATGGCGGGCCGGCTATGATGTTCAACGCACAG 176
Db 1157 AGTACCGCGCTATGTTTGGCGGATGCTTGACAAATGCGGCTACATATCCAGACCAA 1216

Qy 177 TACAAGTTCCTGTGATCCACCGTGAAGTCAATCCCGCGGTGGTGGTCCATACCCAGAA 236
Db 1217 TACCGGCATCTGGGCAATTTTCAAGAAGCACATATCCCTTTTCTTGGTGTCTATCCAA 1276

Qy 237 AAGGTCAGCCCATGCACTGGAAGAGTCACTCACAGCTTCGAGTCTCTCTTCGAGCTG 296
Db 1277 AAAGACAGGAAA---GATGGCTCAGATCTTCCAGATGTCGCTTCTTTGGAAATG 1333

Qy 297 AGCTTCAATTACTCCAAATCACTACTACGGTTTGCATTCGAGCCCTCGGTTCCCTGAGC 356
Db 1334 AGCTTGAATTTGACCGACTCTGTTTTCGATATACATACGAGCCCATCAATGAAGTACG 1393

Qy 357 GGAACGAGAGATGATCAATTCACACCCAGGCAATCAGGCTGTTCTCCAGCACTCAAG 416
Db 1394 GGGACGAGAGAAATCGTTTCAATACATTCGCAATTTATGGCAAGTGTCCAAAACCTGCA 1453

Qy 417 GCATGTTCCAGGGTTGACCTGGAAATGGTTGATCAATTCATTAAGCATTTGCTGCT 476
Db 1454 CAGATTCAAGCGGTATCGACTTGGAGTGGTTTGTAGTTTACTTCAAAGATGAGTGAAGCTG 1513

Qy 477 TCGAGGAGAGAGCTCGGACTCTGTAGTACGAGATTTAGATCCCGCTCTTCAAGACA 536
Db 1514 GACGAATCGGAATCGGCCACACTTCAAGATTAATGAGCTGCTCAAGGAGCAGATAAGAGC 1573

Qy 537 CAGAACAACTCGCAGCCGATCTGGAGCCATCTGGGATATGCTTTGAAGACCTTACATC 596
Db 1574 CAGAACAACTTGGCTTGGATCTCAA---AAGAGCCAGTTTGGCTCTCAAGATTACTTC 1630

Qy 597 TACCGCGGATCAAGTCGATCGGACCGGAGCCCAAAAGAGAGACTCATGTTTGAACGC- 655
Db 1631 TATCCGCATCTTAAATCGATCGGACCGGCAAAATCCACACAGATCTCATCTTCGACTCT 1690

Qy 656 -----AATCAAGCTCGCCAGCAAGTTTGGCAAGTTTGGCACTCCACCTGCAATCTCCGAG 710
Db 1691 GTGTTCAACTGTGCGAGAGCATGACAGCATACAGCCCGCTGTCAGGTATTTGCGGAC 1750

Qy 711 GAGTTT-----ATAGTTCGAGCAGACCCCAACCTCTCTCGGCCAC 749
Db 1751 TATGTTTTCGCGGCAAAATCATTCGCGAATCAGACCAACATAGACCCCTACATCGCGCA 1810

Qy 750 TTTCTCTCATGCGATTTGGTCAAGCCGTCGAGTCCCGAATCAAGTCTACTGTATGGA 809
Db 1811 CTCTTGTATGCGATTTGATCGATCCGCGCAAGTCTCGGCTCAAGATATACCTGCTGAG 1870

Qy 810 CGCCAGCTCGACCTGCGCTCCATCGAAGGATTTTGGACTCTCAACCGGCGCAGCAACGAT 869

```

```

Db 1871 AAGACGGTCTCGTTGTCGGTATGGAAGATCTGTGGACGCTGGGCGGCCAACGAGTCGAC 1930
Qy 870 CCAGAGACACTGGATGGTCTGGATCGCTGAGGAGCTGTGGCAGCTATTGCCCCGTCAAG 929
Db 1931 GCATCCACCATGGACGGCTTGACATGCTTCGCGAGCTCTGGAGCTGCTAAAGATTCCC 1990
Qy 930 GAGGTCTGTCTCCACTCGCGAATGCTTTTACAGCCGGGTACTCTCACGCGAGGAGCAG 989
Db 1991 ACTGCCACTTGGAGTATCCAAAAGGATCTGGAATTTGGGAGAAATTCGCAACGAGCAG 2050
Qy 990 CTCCTCTCATTAATAATTTTACCTTGTCTCTAAAGCGCACTTCCGGAACACACATC 1049
Db 2051 CTCCTCATTCATGGCCACTATACACTGCACCAACAAACCCCATGCTGAGCCCCAGGTG 2110
Qy 1050 TATTTCCCTGCTTTTGGGAGAACGACAAACCAATCCGGAAGGATTGGCCACCTCTTTT 1109
Db 2111 TATTTACGGTTCCTGGCATGATGACGCGGAAATCAGCAATGCTTTGACCATCTTCTTC 2170
Qy 1110 GAGAGCAGAGTTCGGGTGCTTGGCTAAGAGCTATCCAG 1149
Db 2171 CAACGTCAAGATTGAGCATGCGGAAAGTACCGAG 2210

RESULT 4
CPU011963 Claviceps purpurea dl gene. 2182 bp DNA linear PLN 14-FEB-2003
LOCUS Claviceps purpurea dl gene.
DEFINITION Claviceps purpurea dl gene.
ACCESSION AJ011963
VERSION AJ011963.1 GI:4499836
KEYWORDS D1 gene; dimethyl-allyl-tryptophan-synthase.
SOURCE Claviceps purpurea (ergot fungus)
ORGANISM Claviceps purpurea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.

REFERENCE
1 Tudzynski, P., Holter, K., Correia, T., Arntz, C., Grammel, N. and Keller, U.
Evidence for an ergot alkaloid gene cluster in Claviceps purpurea
Mol. Gen. Genet. 261 (1), 133-141 (1999)
99168777
PUBMED 10071219
REFERENCE 2 (bases 1 to 2182)
Tudzynski, P.
Direct Submission
Submitted (13-OCT-1998) Tudzynski P., Westfaelische
Wilhelms-Universitaet, Institut fuer Botanik, Schlossgarten 3,
Muenster, D-48149, GERMANY
Location/Qualifiers
1. 2182
/organism="Claviceps purpurea"
/mol_type="genomic DNA"
/strain="P1"
/db_xref="taxon:5111"
512..1977
/ gene="dl"
join(-512..1677,1743..1864,1919..>1977)
/ gene="dl"
join(512..1677,1743..1864,1919..1977)
/ gene="dl"
/codon_start=1
/product="dimethyl-allyl-tryptophan-synthase"
/protein_id="CA39314.1"
/db_xref="GI:4499837"
/db_xref="SPTREMBL:O94204"
/translation="MSTADKPGNGVYEIILSLIFDPPSQRLLWVHSTAMPFAAMLNDNA
YNIHQYRLGIFKKHIIPLGVPTKDKRWLSILTRCGLPLELSNLTDSVVRYT
YEPINVTGKEDTFLAITSVKLOAIQAGIDLEWFSYKDELTLDESSTATQOS
NELAKQIKTONKALDLKESOPALAKVFEHLKSTATGKSTHDLIFDSVLKLSOKHD
SIQAPQVLCIDYVSRNHSANEVDQGHARLLSCDLIDPAKSRVKIYLLEKTVLSV
MEDLTGGGQVADASTMDGLMRLWLSLLKVPFTHLEITGKGLGEIIFNQLPSNA
NYTLHNDMPPEQVYTFVFGMDASINALTIFQFHGFDDMAKNRYFLQDSYPYH
DFESLNYLHAYISFSYRRNKPYLSVYLHTFETGRWFEVADSPISFDAYRCELSTK"

```

exon <512..1677
/gene="di"
/number=1
intron 1678..1742
/gene="di"
/number=1
exon 1743..1864
/gene="di"
/number=2
intron 1865..1918
/gene="di"
/number=2
exon 1919..>1977
/gene="di"
/number=3

ORIGIN

Query Match 15.1%; Score 210.8; DB 8; Length 2182;
Best Local Similarity 52.2%; Pred. No. 2.7e-51;
Matches 585; Conservative 0; Mismatches 502; Indels 33; Gaps 4;

QY 57 TTCTAGCTGCTGAGTCAAGCCCTGAACCTCTGACACAGGACACACAAAATGGTGTAT 116
DB 545 TAGCAAAATTGAGTCTGATCTTTGATTTCCCGAGCAATGAACAGCACTATGGTGGAC 604
QY 117 AGCAGAGCTCCGATGTTTCCACCATGATGGGGGGGGCGGCTATGATGTTCCAGCACAG 176
DB 605 AGTACCGCGCTATGTTTGGCGGATGCTTGACATGCGGCTACAATATCCACGACCAA 664
QY 177 TACAGTTCTCTGTATCCACCGTGAAGTATCATCCCGCGGCTGGTCCATACCCAGAA 236
DB 665 TATCGGCATCTGGGCATTTTCAAGAGACCAATATCCCTTTCTTGGTGTCTATCCAACA 724
QY 237 AAGGGTCAGCCCATGCACTGGAAGAGTATCTCACAGCTTCGGACTTCCTTTTCGAGTGT 296
DB 725 AAGACAAGMAA---GATGGCTCAGATCCTCACAGATGCGCTTCCTTTGGAATG 781
QY 297 AGCTTCAATCTCAAAATCACTACTACGCTTTCGATTCGAGCCCTCGGTTCCCTGACG 356
DB 782 AGCTTCAATTTGACGACTCTGCTGCTGATATACATACGAGCCCATCAATCAAGTACG 841
QY 357 GGAACGAAGATGATCCATTCACACCCAGGCAATCAGGCTGTCTCCAGACCTCAAG 416
DB 842 GGGACGGAAGATGAGTTCAATATGCGCAATTTATGACAGTGTCCAAAATCTGCA 901
QY 417 GCATGGTTCCAGGGCTTGACCTGGAATGTTTCGATCATTTCACTAAAGCATTTGTCGTT 476
DB 902 CAGATTCAAGCGGTATCGACTTTGAGTGTGTTAGTTACTTCAAGATGAGCTGACGCTG 961
QY 477 TCGAGAGAGAGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCCGTCTTCAGACA 536
DB 962 GACGAATCGGAATCGGCCACACTTCAAGATTAATGAGCTGTCAAGGAGCAGATAAGAGC 1021
QY 537 CAGAAACAACTGGCAGCCGATCTGGAGCCATCTGGCGATATTGTCTTGAAGACCTACATC 596
DB 1022 CAGAACAGTTAGCTTGGATCTCAA---AGAAGCCAGTTGCGGCTCAAGTTTACTTC 1078
QY 597 TACCGCGGATCAAGTTCGATTCGCAACGGGACCCCAAAAGAGAGACTCATGTTGA---C 653
DB 1079 TATCCGCATCTTAAATCGATCGGACCGGCAAAATCCACACGATCTGATCTTTCGACTCC 1138
QY 654 GCAATCAAGCTCCCGCAGATTTGGCAAGTTGCAAGTTGCCACTCCACTGCAATCTCCGAGG 713
DB 1139 GTGTTGAAGCTGTGCGCAGAGCATGACGATACAGCCCGGTTCCAGGTATTGTGCGAC 1198
QY 714 TTTAT-----AGCTGAGCGAGCACCCACCTCTCCGCGCCAC 749
DB 1199 TATGTTTCGGCGGAAATCATCTCCGAGAAGTAGACCAACAGGAGCCCTACATGCGCA 1258
QY 750 TTTCTCTCATGCAATTTGTCAGCGCTCGAGTCCGATCCGATCAAGTCTACTGTATGGA 809
DB 1259 CTCTTGTCATGATTTGATCGATCCGCGCAAGTCTCGGCTCAAGATATACCTCTGGAG 1318

QY 810 CGCCAGCTCGACCTGGCCTCCATCGAAGGTATTGGACTCTCAACGGGCGACGAAACGAT 869
DB 1319 AAGACGGTCTCGTTGTCGGTGTGGAAGATCTGTGGACGCTGGGCGGCGCAACGATCGAC 1378
QY 870 CCAGAGACACTGGATGCTCTGGATGCGCTCAGGAGAGCTGTGGCAGCTATTGCCCGTCAAG 929
DB 1379 GCATCCACCATGGACGGCTTGACATGCTTCGGAGCTCTGGAGCTGTCTGAAAGTTCCC 1438
QY 930 GAGGTCGTGTCCACTGCCGAACTGCTTTTACGAGCCGGGTACCTCACCGCAGGAGCAG 989
DB 1439 ACTGCCCACTTGGAGTATCCAAAGGGTATCTGAATTGGAGAAATTCGAAACGAGCAG 1498
QY 990 CTCCTCTTCATTAATAATTTTACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1049
DB 1499 CTCCCATCCATGGCCAACTATACATCGACACACAAACGACCCATGCTGAGCCCGAGTG 1558
QY 1050 TATTTCCCTGCTTTTGGGAGAGACGACAAAACCATCTCCGGAAGATTTGGCAGCTTTT 1109
DB 1559 TATTTACGGTTTTCGGCATGAATGACGCCGAAATCAGCAATGCTTTGACCATCTTCTTC 1618
QY 1110 GAGAGCAGAGTTGGGCTGGCTTGGCTAAGAGCTATCCAG 1149
DB 1619 CAACGTACGGATTGACGACATGGCGAAAACCTACCGAG 1658

RESULT 5
AY259840
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source

AY259840 2299 bp DNA linear PLN 12-DEC-2003
Claviceps purpurea dimethylallyltryptophan synthase (dmaW) gene,
dmaW-1 allele, complete cds, alternatively translated.
AY259840
AY259840.1 GI:32402653
Claviceps purpurea (ergot fungus)
Claviceps purpurea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
1 (bases 1 to 2299)
The determinant step in ergot alkaloid biosynthesis by a grass
endophyte
Fungal Genet. Biol. (2004) In press
2 (bases 1 to 2299)
Wang, J. and Schardl, C.
Direct Submission
Submitted (17-MAR-2003) Plant Pathology, University of Kentucky,
201 Plant Sciences Bldg., Lexington, KY 40546-0312, USA
1. 2299
/organism="Claviceps purpurea"
/mol_type="Genomic DNA"
/isolate="ATCC20102"
/specific_host="Secale cereale"
/db_xref="ATCC:20102"
/db_xref="taxon:5111"
<600..>2116
/gene="dmaW"
/allele="1"
join(<600..1816,1882..2003,2058..>2116)
/gene="dmaW"
/allele="1"
/product="dimethylallyltryptophan synthase"
join(600..1816,1882..2003,2058..2116)
/gene="dmaW"
/allele="1"
/EC_number="2.5.1.34"
/note="aromatic prenyltransferase; L-tryptophan
dimethylallyl transferase;
4-(gamma,gamma)-dimethylallyltryptophan synthase; DMAT
synthase; 4-(3-methylbut-2-enyl)-L-tryptophan synthase;
dimethylallyl-pyrophosphate; most likely true start codon
might be the short isoform based on comparison with
GenBank Accession number L39640; both long and short

gene
mrna
CDS

isoforms give rise to functional protein as confirmed by heterologous expression in yeast"

/codon_start=1
/evidence=experimental
/product="dimethylallyltryptophan synthase long isoform"
/protein_id="AAP81209.1"
/db_xref="GI:32402654"
/translation="MCETEVDLSQCPKVAAMSTAKDPNGVVEILSLIFDPPSNEOR
LWHSRAPAPAMLDNAGYNIDYVHGLIFKXHLIPPLGVVPTDKRMLSLRQCG
LPLESLNCDTSVVRITYEPINEVTGETDNTNLAIMTSVKLQAIQGLWFPSY
KDELILNCDTSVVRITYEPINEVTGETDNTNLAIMTSVKLQAIQGLWFPSY
THDLIFDSVLKLSQKHSIQAPFVLQVCDYVSRNHSAEVDQGHARLLSCLDLPDA
KSRVLYLEKTVLSLWMDLMTLGGQVDSRMDGLMLRELMLSLKLVPTGHLHPK
GYLELGEIPNEQPSNANTYLNHNDMPPEPVFTVFGMNDAEISNALTIFPQHGFD
DMAKNRVFLQDSVPYVHDPESLNLYHAYISFSYRNKPYLSVLHTFETGRVPFVADS
PISFDAYRRCLESTK"
join(651..1816,1882..2003,2058..2116)
/gene="dmaW"
/allele="1"

CDS

/note="aromatic prenyltransferase; L-tryptophan dimethylallyl transferase;
4-(gamma,gamma)-dimethylallyltryptophan synthase; DMAT synthase; 4-(3-methylbut-2-enyl)-L-tryptophan synthase; dimethylallyl-pyrophosphate; most likely true start codon might be the short isoform based on comparison with GenBank accession number L39640; both long and short isoforms give rise to functional protein as confirmed by heterologous expression in yeast"

/codon_start=1
/evidence=experimental
/product="dimethylallyltryptophan synthase short isoform"
/protein_id="AAP81210.1"
/db_xref="GI:32402655"
/translation="NSTAKDPNGVVEILSLIFDPPSNEORLWHSRAPAMLDNAGYNIDYVHGLIFKXHLIPPLGVVPTDKRMLSLRQCG
LPLESLNCDTSVVRITYEPINEVTGETDNTNLAIMTSVKLQAIQGLWFPSY
KDELILNCDTSVVRITYEPINEVTGETDNTNLAIMTSVKLQAIQGLWFPSY
THDLIFDSVLKLSQKHSIQAPFVLQVCDYVSRNHSAEVDQGHARLLSCLDLPDA
KSRVLYLEKTVLSLWMDLMTLGGQVDSRMDGLMLRELMLSLKLVPTGHLHPK
GYLELGEIPNEQPSNANTYLNHNDMPPEPVFTVFGMNDAEISNALTIFPQHGFD
DMAKNRVFLQDSVPYVHDPESLNLYHAYISFSYRNKPYLSVLHTFETGRVPFVADS
PISFDAYRRCLESTK"

ORIGIN

Query Match 15.1%; Score 210.8; DB 8; Length 2299;
Best Local Similarity 52.2%; Pred. No. 2.7e-51;
Matches 585; Conservative 0; Mismatches 502; Indels 33; Gaps 4;

QY 57 TTCTAGCTGTGAGTCAAGCCCTGGAACCTCTCGAACAGGACACACAAAATGGTGGTAT 116
DB 684 TACGAAATTTGAGTCTGATCTTGATTTCCCGAGCAATGAACAGGACTATGGTGGCAC 743

QY 117 AGCAGAGCTCCGATGTTTGGCCACATGATGGCGGGGGCGGCTATGATGTTCAAGCAG 176
DB 744 AGTACCGCGCTATGTTTGGCGGAGTGTGACAAATGCGCGGTACAATATCCAGACCAA 803

QY 177 TACAAGTCTCTGTATCCACCGTGAAGTGCATATCCCGCGGTGGGTCCATACCACGAA 236
DB 804 TATCGCATCTGGCATTTTCAAGAGACATATNCCCTTTCTTGGTGTCTATCAACA 863

QY 237 AAGGTGAGCCCATGCACTGGAAGAGTCACTACACGCTTCGGACTTCCTTTCAGCTG 296
DB 864 AAAGACAAGGAA--GATGGGTCAAGTCTTCAATATGGCAATATGACAAAGTGTCCAAACCTGCA 920

QY 297 AGCTTCAATTTACTCCAAATCACTACTACGTTTGCATTCGAGCCCTCGTTTCCCTGAGC 356
DB 921 AGCTTGAATTTACCGACTCTGCTGTTTCAATATACATACAGCCCATCAATGAAGTGA 980

QY 357 GGACGAGAGATGATCCATCAACCCAGGCAATCAGCCCTGTTTCTCAGGACCTCAAG 416
DB 981 GGACGAGAGAGATACGTTTCAATATGGCAATATGACAAAGTGTCCAAACCTGCA 1040

QY 417 GCCATGTTCCAGGGCTTCACTCGGATGTTGCTGATCTTCACTTAAGCATGTTGCTGCT 476

DB 1041 CAGATTCAAGCGGTATGAGCTTGGAGTGGTTTACTTTCAAGATGAGCTGACGCTG 1100
QY 477 TCGAGGAGAGGCTCGGACTCTGCTAGATCGAGATATTGAGATCCCGCTTCTCAAGACA 536
DB 1101 GACGAATCGGAATCGGCACACACTTCAAAAGTAATGAGCTGTGCTCAAGGACGAGATAAGACG 1160
QY 537 CAGAAACAACTGGCAGCGGATCTGGAGCCATCTGGCCATATTGCTTTGAAGACCTACATC 596
DB 1161 CAGAACAGTTAGCTTGGATCTCAA--AGAAAGCCAGTTGGCGTCAAGATTACTTC 1217
QY 597 TACCCGGGATCAAGTCTGAGTCGGACCGGACCCAAAAGAGAGATCATCTGTTTGA--C 653
DB 1218 TATCCGATCTTAAATCGATCGGACCGGCAAAATCCACACAGCATCTGATCTTCGACTCC 1277
QY 654 GCAATCAAGGCTGCGCAAGATTGGCAAAAGTTGCCACTCCACTGGCAATCTCTCGAGGAG 713
DB 1278 GTGTTGAAGCTGTCGAGAGCATGACGATACAGCCCGCTTCAGGTTATTGTGGCAG 1337
QY 714 TTTAT-----AGCTGAGGAGCACCCACCTCTCTCGGCCAC 749
DB 1338 TATGTTTCGGCGGAAATCATTCGCGAAGTAGACCAACACGAGGCCCTACATGCGCGA 1397
QY 750 TTTCTCTCATCGGATTTGGTCAAGCGCTCGAGTCCCGAATCAAGTCTACTGTATGGAA 809
DB 1398 CTCTGTCTGCGGATTTGATCGATCCCGCAAGTCTCGCGTCAAGATATACCTACTGGAG 1457
QY 810 CGCAGCTCGACCTGGCTCCATCGAAGGTATTGGAATCTCAACGGGCGACGGAACGAT 869
DB 1458 AAGACGGTCTCGTTGTCGGTATGGAAGATCTGTGGACGCTGGGCGGCCAACAGTCGAC 1517
QY 870 CCAGAGACATGATGTTGATGCGCTGAGGAGCTGTGGAGCTATTGCCCCGTCAAG 929
DB 1518 GCATCCACCATGAGCGGCTTGACATGCTTCGGGAGCTCTGGAGCCCTGTGAAAGTCCC 1577
QY 930 GAGGCTCTGTCTCCACTCGCCAACTGCTTTTACGAGCGCGGTACCTCACCGCAGGAGCAG 989
DB 1578 ACTGGCCACTTGGAGTATCCAAAGGATCTCGAATTTGGAGAAATTCGAGACGAGCAG 1637
QY 990 CTCCTCTTCAATATAATTTTACCTGTCTCTTAAAGCGCACTTCCCGAACACAGATC 1049
DB 1638 CTCCTATCCATGCGCAACTATACACTGCACCAACAGACCCCATGCTGAGGCCCGAGTG 1697
QY 1050 TATTTCCCTGCTTTTGGGAGAGACGACAAACCATCGCGAAGGATTTGGCCACCTTCTTT 1109
DB 1698 TATTTTACGTTTTCGGCATGATGACGCGAATATAGCAATGCTTTGACCATCTTCTTC 1757
QY 1110 GAGAGCAGAGTTGGGTGGCTTGGCTTAAGAGCTATCCAG 1149
DB 1758 CAACGTCACGATTTGACGACATGGCGGAAAACTACCGAG 1797

RESULT 6

CPU312753

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

CPU312753 2694 bp DNA linear PLN 24-APR-2001
Claviceps purpurea cpd2 gene for putative
dimethyl-allyl-tryptophan-synthase, exons 1-3, strain T5.
AJ312753
AJ312753.1 GI:13810197
cpd2 gene; dimethyl-allyl-tryptophan-synthase.
Claviceps purpurea (ergot fungus)
Claviceps purpurea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
1
Arntz, C. and Tudzynski, P.
Molecular analysis of dimethyl-allyl-tryptophan-synthase genes
Unpublished
2 (bases 1 to 2694)
Tudzynski, P.
Direct Submission
Submitted (20-APR-2001) Tudzynski P., Institut fuer Botanik,
Westfaelische Wilhelms-Universitaet, Schlossgarten 3, 48149
Muenster, GERMANY

FEATURES	Location/Qualifiers	
source	1..2694	
	/organism="Claviceps purpurea"	
	/mol_type="genomic DNA"	
	/strain="T5"	
	/db_xref="taxon:5111"	
gene	901..2366	
	/gene="cpd2"	
CDS	join(901..2066,2132..2253,2308..2366)	
	/function="isoprenylation of tryptophane"	
	/codon_start=1	
	/product="putative dimethyl-allyl-tryptophan-synthase"	
	/protein_id="CAC37396.1"	
	/db_xref="GI:13810198"	
	/db_xref="SPTREMBL:Q9C141"	
	/translation="MSTAKDPNGVYVEILSLIFDPFNEQRLWHSHTAPMFAALMDNA GYSVDYVHLSGIFKTHIIPFLGVPTKGOERWLSILTRCGLELSLNCDSVVRVA YPIINMTCTEDPSNLTPIIGSVOKLQIOAGIDLEWFSYFKDLITDESEAILQD TLVKEQIITQNKLALDLKESQFALKVFPYPLKSIATGNSHFLIFSVFKLSQKHD SICPAQALCDVSRNDSSEVDQERHALHRLSCDLIDPAKSRVKIYLQSGTVSLPA MEDMTLGGRRVDASTMDGLMRELWLLKVTGHELYPKGYMELGIPNEQPSLV NYTLHNDMPFPQVYFTVFGNDABISNALTIFLRHGFADMAKRYVFLQDSYPYH DFESLNLHLSVFSFRRNKPYLSVYLHTFETGRWPVWADSPISFDAYRRCDLSFK"	
exon	901..2066	
	/gene="cpd2"	
	/number=1	
intron	2067..2131	
	/gene="cpd2"	
	/number=1	
exon	2132..2253	
	/gene="cpd2"	
	/number=2	
intron	2254..2307	
	/gene="cpd2"	
	/number=2	
exon	2308..2366	
	/gene="cpd2"	
	/number=3	
ORIGIN		
Query Match	15.1%; Score 210.6; DB 8; Length 2694;	
Best Local Similarity	52.8%; Pred. No. 3.2e-51;	
Matches	590; Conservative 0; Mismatches 494; Indels 33; Gaps 5;	
QY	57	TTCTAGTGTGAGTCAAGCCCTGAACCTCTGAAACAGGACACACAAATGTTGGTAT 116
Db	934	TACGAATTTTGTGCTGATTTTGTATTTCCCGAGCATGAAACAGGACTATGGTGGAC 993
QY	117	AGCACAGCTCCGATGTTGCCACCATGATGGCGGGCCGGCTATGATGTTCCAGCACAG 176
Db	994	AGTACGGCGCTATGTTGGCGGATGCTTGACAAATGCCGGCTACAGTGTCCACGACAA 1053
QY	177	TACAAGTTCCTGTATCCACCGTGAGGTGATCATCCCGCGGTGGGTCCATACCCAGAA 236
Db	1054	TATCGGCATCTGAGCATTTTCAAGACACATATATCCCTTTCTTGGTGTCTATCCACA 1113
QY	237	AAGGGTCAGCCCATGCACTGGAGAGTATCTCACACGCTTCGGACTTCCTTTCCGAGTGT 296
Db	1114	AAAGGCCAGGAA--GATGGCTCAGCATCTCCACAGATGTTCTTCTTTGGACTG 1170
QY	297	AGTTCAATCTCCAAATCACTACTACGTTTGGATTCGAGCCCTCGGTTCCCTGAGC 356
Db	1171	AGCTTAACTGTACCGACTCTGTGTTGATAGCATACGAGCCCATCAATGAGATGAGC 1230
QY	357	GGAAACGAAGATGATCCATTCACACCCAGGCAATCAGGCCCTGTTCTCCAGGACCTCAAG 416
Db	1231	GGACCGGAGAAAGATCCATCCATACGTTGCCGATTTAGGAGTGTCCAAAGCTTGGC 1290
QY	417	GCAATGGTCCAGGGGTTGACCTGGAAATGTTGATCATTTCACTAAAGCATTTGGTGT 476
Db	1291	CAGATTCAAGCGGGTATCGACTTGGAGTGTTTCTTCTTCAAGGATGAGTTGACGTTG 1350

QY	477	TCGAGGAAGAGGCTCGGACTTGTCTAGATCGAGATATTGAGATCCCCTGCTTCAAGACA 536
Db	1351	GACGAATCTGAATCGGCATACCTTCAAGATAGTACTGAGCTGGTCAAGAGGACGATAAGACG 1410
QY	537	CAGAACAACTGGCAGCCGATCTGGAGCCATCTGGCGATATTGCTTTGAAGACCTACATC 596
Db	1411	CAGAACAACTTAGCTTGGATCTCAA---AGAAAGCCAGTTCGCGCTCAAAAGTTTACTTC 1467
QY	597	TACCCGCGGATCAAGTCGATCGGACCGGACCCCAAAAGAGAGACTCATGTTTGAC--- 653
Db	1468	TATCCGATCTCAAAATCGATCGGACCGGCAATTTCCACACACTTTTCTCATCTTCGACTCT 1527
QY	654	GCAATCAGGCTGCGGACAAAGTTG---GCAAGTTGCGCACTCCACTGCGCAATCCTCGAG 710
Db	1528	GTGTTCAAGTGTGCGAAGCATGACAGCATACAGCCCGGTTTCCAGGATTTGTGCGAC 1587
QY	711	GAGTTT-----ATAGTCTGAGGAGGACCCACCCCTCTCTCGGCCAC 749
Db	1588	TATGTTTCGGCGGAAATGATTCCTCAGAAAGTAGACCAACACAGAGCCCTACATGCGCA 1647
QY	750	TTTCTCTCATCGGATTTGGTCAAGCCGTCGAGTCCCGAATCAAGTCTACTGTATCGAA 809
Db	1648	CTCTTGTATCGGATTTGATCGATCCCGCAAGTCTCGCGTCAAGATATACCTGACGAG 1707
QY	810	CGCCAGCTCGACCTGGCTCCATCGAAGTATTGGAGTCTCTCAACGGCGGACGAAACGAT 869
Db	1708	CAGACGGTCTCATTTGCCGCGATGGAAGATCTGTGACGCTGGCGCGCGACGAGTCGAC 1767
QY	870	CCAGAGACACTGGATGTTGGATGCGCTGAGGAGTGTGGCAGCTATTGCCCGTCAAG 929
Db	1768	GCATCCACCATGGACGGCTTGACATGCTTCGCGAGCTCTGGAGCCTGCTGAAAGTTCCC 1827
QY	930	GAGGTCGTGTCCTCCTCCGAACTGCTTTTACGAGCGGGTACTCTCACCGCAGGACGAG 989
Db	1828	ACTGCCACTTGGAGTATCCAAAGGATATATGGAATTTGGAGAAATTCCTTAACGAGAG 1887
QY	990	CTCCCTTTCATATAAAATTTTACCTTGTCTCTTCTTAAAGCGCACTTCCCGAACCAAGATC 1049
Db	1888	CTTCCATCCCTGGTCAACTACACACTGCACGCAACGACCCCATGCTGAACCCCGAGTG 1947
QY	1050	TATTTCCCTGCTTTTGGCAGACGACAAACCATTCGCGAAGATTGGCCACCTTTT 1109
Db	1948	TATTTCAAGTTCGCGCATGATGACGCGAAATCAGCAATGCTTTTGACCAATTTTCTC 2007
QY	1110	GAGACGAGATTTGGGTGGCTTGGCTTAAGAGCTATC 1146
Db	2008	CAAGTCAAGTTCGCGCATGCGGCAAAAGTACC 2044
RESULT 7		
AR181512		
LOCUS	AR181512	1347 bp DNA linear PAT 20-APR-2002
DEFINITION	Sequence 1 from patent US 6335188.	
ACCESSION	AR181512	
VERSION	AR181512.1	GI:20223726
KEYWORDS	Source	Unknown.
ORGANISM	Unknown.	
REFERENCE	1 (bases 1 to 1347)	Unclassified.
AUTHORS	Schardl, C.L. and Wang, J.	
TITLE	Endophyte ergot alkaloid synthetic compounds, compounds which	
JOURNAL	encode therefor and related methods	
FEATURES	patent: US 6335188-A 1 01-JAN-2002;	
source	Location/Qualifiers	
	1..1347	
	/organism="unknown"	
	/mol_type="unassigned DNA"	
ORIGIN		
Query Match	14.7%; Score 205.2; DB 6; Length 1347;	
Best Local Similarity	50.8%; Pred. No. 1.2e-49;	
Matches	647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;	

QY 33 GCAGCAACACTGCTGCGCAAGCCCTTCTAGCTGCTGAGTCAAGCCCTGACCTCTCGAAC 92
Db |||||
10 GCAAGACACTCCACAGGAGTTTATCATACCTTTAGCGAAACATTTGACTTTGCCAAC 69
QY 93 AAGAACACACAAATAGTGTGTATAGCAGCTCCGATGTTTCCACCAATGATGCGGGG 152
Db |||||
70 AATGACAGAGGCTATGTTGGTGCACAGCAGCGGCCAAATGTTGAAAGATGCTCCAAACT 129
QY 153 GCGGCTATGATGTTACGCGACAGTCAAGTTCCTCTGTATCCACCGTGAGGTCAATC 212
Db |||||
130 GCTAACTATAGCTATGCTGATCGATCGCATCTGGCATTTTAAAGAGCATGTCAT 189
QY 213 CCGGCGTTGGGTCCATACCCAGAAAGGTCAGCCCATGCACTGGAAGAGTCACTCACA 272
Db |||||
190 CTTTTCCTGTTGCTATCTCTTACCAAGAGTGG---CGAGCGGTGGGTAGCATTTCTAGC 246
QY 273 CGTTCGCACTCTCTTTTCGAGCTGAGCTTCAATTAATCTCCAAATCACTACTACCGTTTGA 332
Db |||||
247 AGATACGGAACCCGTTTGAGCTAAGTCTTAATGCTCTGACTCCATAGTTTGGTATACA 306
QY 333 TTGAGCCCTCGGTTCCCTGACGGGAAACGAAAGATATCCATTCACACCCAGGCAATC 392
Db |||||
307 TACGAGCTATTAAACGCGCAACTGGCAGCCATCTGGATCCGTTTCAACACTTTTCGCTATC 366
QY 393 AGGCTGCTCTCAGGAGCCCTCAAGGCCATGTTTCCAGGGCTTCACCTGGAATGTTTCGAT 452
Db |||||
367 TGGAGGCTTAAAGAGCTTATAGATTCACCGCAGGATAGACCTTCAATGGTTTTC 426
QY 453 CATTTCACTAAAGCATTTGGTCTGTTTCGGAGGAAGAGCTCGACTCTGTAGATCGAGAT 512
Db |||||
427 TACTTTAAACAGAGCTTACCTTGACGCAACAGCTCCACGTACCTGCACTCGCAAAAC 486
QY 513 ATTGAGATCCCGCTTTCAAGACACAGAACAACTGGACCGCATCTGGAGCCATCTGGC 572
Db |||||
487 TTGGTTAAGGAACAGATCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACT 543
QY 573 GATATTGCTTTGAAGACTTACATCTACCGCGGATCAAGTCGATCCGACCCGCGGCCCA 632
Db |||||
544 AAGTTCGTACTGAGACCTTACATCTACCGGATGAGTTCGTCGCACTCTGTAATCG 603
QY 633 AAGAGAGACTCATGTTTGAAGCAAT---CAAGGCTGCGGACAAAGTTTGGCAAAAGTTGCC 689
Db |||||
604 GTCCAGGAGCTGCTGTTTGGCTCCGTCGAGCTAGCGCAGAACACAAAGAGTATCCGT 663
QY 690 ACTCCAGTGCATCTCGAGGAGTTAT---AGCTGAGCGAGCACCCACCTCTCGG--- 745
Db |||||
664 CTTGCTTTGAAATGCTAGAAGACTATGTCAGTCTCGCAATAAATTTCTTACCAGGAT 723
QY 746 -----CCACTTTCTCTCATGCGATTGCTTGAAGCGCTCC 779
Db |||||
724 GACAGTCACAATACTCTATTATCTTACGCTTCTCTTTCGACCTGATAAGTCTTACC 783
QY 780 GAGTCCGGAATCAAGTCTACTGTATGGAACGCGAGCTCGACTGCGCTCTCATCGAAGT 839
Db |||||
784 AAGTCTGCTGTAAGATCTACCTCTCGGAACGAATGGTCTGTTGCGCAGCGATGGAAGAT 843
QY 840 ATTTGAGCTCTCAACGCGGACCGGAACGATCCAGAGACACTGATGGTCTGATGCGGTG 899
Db |||||
844 CTTTGGACGCTTGGCGCGCTCGAGAAGATCAGTCCACTATTGAGGATTTGAGATGATC 903
QY 900 AGGAGCTGTGGAGCTATTTGCCGTACAGGAGGCTGTGTCTGCTCACTGCGCACTGCTTT 959
Db |||||
904 CGAGAACTTTGGGGTCTCTCTCAACATGCTCTCTGTTTTCGCGCCCTTACCCTGAGCCTTAC 963
QY 960 TACGAGCGGGTACCTCACCGAGGAGCAGCTCCCTTCAATTATAATTTTACCTTTGCT 1019
Db |||||
964 TTGCGCTCTGGCGCCATTCCCAATGAGCAACTTCGTCATGGCCNAATTACACCTTACAC 1023
QY 1020 CTTAAAGCGCACTTCCGAAACACAGATCTATTTCCTCTGCTTTTGGGAGAACGACAAA 1079
Db |||||
1024 CATAATGATCCCATACAGAACCGCAAGTGTACTTTACTGTGTGGCATGAATGATATG 1083

QY 1080 ACCATCGGGAAGATTGGCCACCTTCTTTGAGACGAGAGTGTGGGTGCTTTGGCTAAG 1139
Db |||||
1084 GAGGTGACTAATGCACTCAGACATCTTTCATGAGGCATGAATGGAGCGATATGGCAAGT 1143
QY 1140 AGCTATCCAGCGATTTGGCATCTTACTATCCGATGCGACCTGCGAGCGCAATCAC 1199
Db |||||
1144 AATACAAAGCCTCCCTCAGGAACTTTTCCGCACTCATGATTACGAAGCCCTGAATAT 1203
QY 1200 CTGAGCGCTGATCTCTCTCTTACAAGGGGAAACCCGTACATGATGATGTGTACCTC 1259
Db |||||
1204 ATCCACTCGTACATTTCTTCTCTCTCCGAAAGAACCAATATTTAAGTGTGTATCTC 1263
QY 1260 CATACCTTCGAGC 1273
Db |||||
1264 CACTCCTTTGAAC 1277

RESULT 8
AR181513
LOCUS AR181513 1353 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 3 from patent US 6335188.
ACCESSION AR181513
VERSION AR181513.1 GI:20223727
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1353)
AUTHORS Schardl, C.L. and Wang, J.
TITLE Endophyte exort alkaloid synthetic compounds, compounds which
encode thefor and related methods
Patent: US 6335188-A 3 01-JAN-2002;
JOURNAL Location/Qualifiers
FEATURES
source 1..1353
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 14.7%; Score 205.2; DB 6; Length 1353;
Best Local Similarity 50.8%; Pred. No. 1.2e-49;
Matches 647; Conservative 0; Mismatches 598; Indels 39; Gaps 5;
QY 33 GCAGCAACACTGCTGCGCAAGCCCTTCTAGCTGCTGAGTCAAGCCCTGACCTCTCGAAC 92
Db |||||
10 GCAAGACACTCCACAGGAGTTTATCATACCTTTAGCGAAACATTTGACTTTGCCAAC 69
QY 93 AAGAACACACAAATAGTGTGTATAGCAGCTCCGATGTTTCCACCAATGATGCGGGG 152
Db |||||
70 AATGACAGGCTATGTTGGTGCACAGCAGCGGCCAAATGTTCCAAAGATGATCTCCAAACT 129
QY 153 GCGGCTATGATGTTACGCGACAGTCAAGTTCCTCTGTATCCACCGTGAGGTCAATC 212
Db |||||
130 GCTAACTATAGCTATGCTGATCGATCGCATCTGGCATTTTAAAGAGCATGTCAT 189
QY 213 CCGGCGTTGGGTCCATACCCAGAAAGGTCAGCCCATGCACTGGAAGAGTCACTCACA 272
Db |||||
190 CTTTTCCTGTTGCTATCTCTTACCAAGAGTGG---CGAGCGGTGGGTAGCATTTCTAGC 246
QY 273 CGTTCGCACTCTCTTTTCGAGCTGAGCTTCAATTAATCTCCAAATCACTACTACCGTTTGA 332
Db |||||
247 AGATACGGAACCCGTTTGAGCTAAGTCTTAATGCTCTGACTCCATAGTTTGGTATACA 306
QY 333 TTGAGCCCTCGGTTCCCTGACGGGAAACGAAAGATATCCATTCACACCCAGGCAATC 392
Db |||||
307 TACGAGCTATTAAACGCGCAACTGGCAGCCATCTGGATCCGTTTCAACACTTTTCGCTATC 366
QY 393 AGGCTGCTCTCAGGAGCCCTCAAGGCCATGTTTCCAGGGCTTCACCTGGAATGTTTCGAT 452
Db |||||
367 TGGAGGCTTAAAGAGCTTATAGATTCACCGCAGGATAGACCTTCAATGGTTTTC 426
QY 453 CATTTCACTAAAGCATTTGGTCTGTTTCGGAGGAAGAGCTCGACTCTGTAGATCGAGAT 512
Db |||||
427 TACTTTAAACAGAGCTTACCTTGACGCAACAGCTCCACGTACCTGCACTCGCAAAAC 486

QY 513 ATTGAGATCCCGTCTTCAAGACACAGCAAACTGGCAGCCGATCTGGAGCCATCTGGC 572
Db 487 TTGGTCAAGAAAGAGATCAAACTCAAAACAACTAGCGTTAGACCTTAA--AGGTGAC 543
QY 573 GATATTGTCTGTAAGACCTACATCTACCCGCGGATCAAGTTCGATCCGACCGGAGCCCA 632
Db 544 AAGTTCGTACTCAAGACCTACATCTACCCGGAATTGAAGTCCGTCGCAACTGTGTAATCG 603
QY 633 AAAGAGAGACTCATGTTGACGCAAT---CAAGCTCCGCAACAGTTTGCAAGTTGCC 689
Db 604 GTCCAGAGCTTGTTGGCTCCGTCGCAAGTACGCGAAGCAAGCATGAAGATATCCGT 663
QY 690 ACTCCACTGCGCAATCTCTGAGGAGTTTAT--AGCTGAGCGAGCAGCCACCCCTCTCGG-- 745
Db 664 CCGTCTTTGAAATGCTAGAAGACTATGTCAGTCTCGCAATAAAGTCCCTACCCAGGAT 723
QY 746 -----CAGTTTCTCTATGCGATTTGGTCAAGCGTCC 779
Db 724 GACAGTCACAATACTCCATTATCTTACGCGCTTCTCTTGGACCTGGTGAATCTTACC 783
QY 780 GAGTCCCGGAATCAAGTCTACTGTATGGAAGCGCAGCTCGACTGCGCTCCATCGAAGT 839
Db 784 AAGTCTCGTCAAGATCTACCTCTGGAACGATGCTGTTGCGCAGCGATGAAGAT 843
QY 840 ATTTGAGCTCTCAACCGGCGACGGAACGATCCAGACACTGGATGTTGGATCGGCTG 899
Db 844 CTTTGGACGCTTGGCGCGCTCGAGAAGATCAGTCCACTATTGAGGGAATTGGAGATGATC 903
QY 900 AGSGAGCTGCGAGTATTGCCGTCACGAGGCTCTGTCTCCACTGCGGACTGCTTT 959
Db 904 CGAGAACTTTGGGTCTCTTAACATGCTCTCGTGTGTCGCGCTTACCTGAGCCTTAC 963
QY 960 TAGAGCCGGGTACCTCAGCGAGGAGCTCCCTTCAATTATAAATTTTACCTTGTCT 1019
Db 964 TTGCCCTCGGCGCATTCCTCAATGAGCACTTCGCTCCATGCGCAATTACACCTTAC 1023
QY 1020 CTTAAAGCGCACTTCCGACACAGATCTATTCCCTGCTTTTGGGAGAGACAGCAAA 1079
Db 1024 CATATGATCCGATACCAAGCGCAAGTGTACTTCTGTTGCGGCAATGAATGATG 1083
QY 1080 ACCATCGCGAAGATTGGCCACTTCTTTGAGACAGAGGTTGGGTGGCTTGGCTAAG 1139
Db 1084 GAGTGACTAATGCACTACAGAAATCTTCATGAGGCATGAATGAGCATATGCAAGT 1143
QY 1140 AGCTATCCAGGATTTGGATCTCTACTATCCGATGTCGACCTGACAGCCCAATCAC 1199
Db 1144 AAATACAAAGCTGCTTGGGAATCTTTCCGCGCATATAATTAAGAGCCCTTAAATTAT 1203
QY 1200 CTGACGGCTGGATCTCTTCTTCAAGGGGAAACCGTACATGATGTTGTAAGTCTC 1259
Db 1204 ATCCACTGTACATTTCTTCTCTTCCGAAATACAGCCCATATTTAAGTGTGATCTC 1263
QY 1260 CATACCTTCAAGC 1273
Db 1264 CACTCATTTGAAAC 1277

RESULT 9
AY262013 2143 bp DNA linear PLN 12-DEC-2003
LOCUS Balansia oblecta dimethylallyltryptophan synthase (dmaW) gene,
DEFINITION dmaW-1 allele, complete cds.
ACCESSION AY262013
VERSION AY262013.1 GI:32967581
KEYWORDS
SOURCE Balansia oblecta
ORGANISM Balansia oblecta
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Balansia.
REFERENCE 1 (bases 1 to 2143)
AUTHORS Wang, J., Machado, C. and Schardl, C.L.
TITLE The determinant step in ergot alkaloid biosynthesis by a grass

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

endophyte
Fungal Genet. Biol. (2004) In press
2 (bases 1 to 2143)
Wang, J., Machado, C. and Schardl, C.L.
Direct Submission
Submitted (24-MAR-2003) Plant Pathology, University of Kentucky,
201 Plant Science Building, Lexington, KY 40546-0312, USA
Location/Qualifiers
1..2143
/organism="Balansia oblecta"
/mol_type="genomic DNA"
/isolate="B249"
/specific_host="Cenchrus echinatus"
/db_xref="taxon:40611"
/tissue_type="sclerotial mycelium"
<340..>1852
/gene="dmaW"
/allele="1"
/gene="dmaW"
/allele="1"
/product="dimethylallyltryptophan synthase"
join(340..1505,1565..1686,1743..>1852)
/gene="dmaW"
/allele="1"
/EC_number="2.5.1.34"
/note="aromatic prenyltransferase;
4-(gamma,gamma)-dimethylallyltryptophan synthase; DMAP
synthase; 4-(3-methylbut-2-enyl)-L-tryptophan synthase;
dimethylallyl-pyrophosphate: L-tryptophan dimethylallyl
transferase"
/codon_start=1
/product="dimethylallyltryptophan synthase"
/protein_id="AAP92451.1"
/db_xref="GI:32967582"
/translation="MATENGPKDSVYETLSLINFPPDDTQRLWHSAPMLAEMLCQC
EYVHNYQQLGIFPKHIIIPYLPYFKSKRWLSILIRIGFIDPELINCDSVVRIT
YEPINEATGTEKOPNTLAIDSAQELIRIOAGIDLEWFTFKDELINMTSDSYLRIS
NLVNCQIKTKNKLALDKGDFTLKVYMPPELKSATGSKHSLIFGVRRLSKYN
SIRAPLMDLDYVTSRNGANGBEHLPLERLSLCDLMDPKHVKRLYHVKVLSGA
MEDLTGGRRTDSTTMDGLNMRDLWLEITCLOKYAPFLELGOIPEOLPSMA
NYTLHGDPMPEQVYFTVGMNDSKVISALTREKERVNGMKRYAFLQNSYND
DHESLYLHTYIISFYRKNPKYLSVILHTFESSGWPFPDSSATFTYRRCDLSCKDN
GLDDKLSPVQCQRV"

ORIGIN

Query Match 13.9%; Score 194; DB 8; Length 2143;
Best Local Similarity 51.1%; Pred. No. 2.9e-46;
Matches 582; Conservative 0; Mismatches 525; Indels 33; Gaps 4;
QY 56 CTTCTAGCTGCTGAGTCAAGCCCTGAACTCTCGAACAGGACCAACAAATGGTGGTA 115
Db 372 CTACGAGACTCTCAGTCTCATTTTCAACTTCCAGACGATACGCAAGACTGTGTGGCA 431
QY 116 TAGCACAGCTCCGATGTTGCCACCATGATGGCGGGCGGCTATGATGTTACGCA 175
Db 432 CAGTACCGCAACATGCTTGAGAAATGCTTCAGACCTGGAATACAGTGTTCATACCA 491
QY 176 GTACAAATTCCTGTATCCACCGTGAGGTCACTATCCCGCGGTGGGTCCATACCCAGA 235
Db 492 ATATCAACAGCTAGGTATCTTCAAAAAGCACATAATTCCTATTGGGTGTTTATCCAC 551
QY 236 AAGGGTCAGCCCATGCACTGGAAGACTCATCTCACAGCTTCGGACTTCCTTTCGAGCT 295
Db 552 AAAGAGTAAGAAAGA---TGGGTGACATTTTGACAAAGATATGGGATTCCTTCGAACT 608
QY 296 GAGCTTCAATTAATCTCCAAATCACTACTACGTTTGGATTCGAGCCCTCGGTTCCTGTAC 355
Db 609 GAGCTTGAATGCTCCGACTCCGCTCGTTCGATACATACGACCACTTAATGAGGCAAC 668
QY 356 GGGAAACAGATGATCCATTCAACCCAGGCAATCAGGCTGTTTCCAGAGACTCAA 415
Db 669 GGGACCGGAAAGATCCATACACATTTGGCGATTTCTTGACAGTGGCAGGAATAT 728

Db 1361 ATCGGGATCGGCCATCTTCAAGATACTGAGCTGGTCAAGGACGACATMAAGACGAGAA 1420
Qy 542 CAAACTGGCAGCCGATCTGGAGCCATCTGGCGATATTTGTTTGAAGACCTTACATCTACCC 601
Db 1421 CAACTAGCCTTGGATCTCAA---AGAACCCAGTTTCGGCTCAAAGTTTACTTCTATCC 1477
Qy 602 GCGGATCAAGTCGATCGCGACCGGACCCCAAAGAGAGACTCTATGTTTACGC----- 655
Db 1478 GCATCTCAATTCGATCGCGACCGGCAATCCACACACTTTCTCATCTTTGACTCCGTGTT 1537
Qy 656 -----AATCAAGGCTGCCGACAAAGTTTGGCAAGTTTGGCAAGTTTGGCACTCCACTGGCAA 702
Db 1538 CAACTTGTGCGAAGCATGACAGCATACAGCCGGTTCACAGCATTTGGGACTATGT 1597
Qy 703 TCTCGAGGAGTTTATAGCTAGGAGCACCC-----CCCTCTCGGCCACTTTCT 754
Db 1598 TTCCGCGCGAAATTAATTCGCGATGTAGACCAACACGAGCCCTACATCGCGCTCTTT 1657
Qy 755 CTATGCGATTTGTCGAGCGCTCCGAGTCCCGAATCAAGGTCTACTGTATGGAAGCCCA 814
Db 1658 GTATGCGATTTGATCGATCCCGCAAGTCTCGGTCAAGATATACCTGAGAGGAGAC 1717
Qy 815 GCTCGACCTGGCTCCATCGAAGTATTTGGATCTCAACGGCGAGCGGAACGATCCAGA 874
Db 1718 GGTCTCATTTGTCGCGATGGAAGATCTGTGAGCGCTGGGCGGCGCAAGTCCGACGATC 1777
Qy 875 GACACTGGATGTTGATCGCTGAGGAGCTGTGCGAGCTATTGCCCGTCAAGGAGG 934
Db 1778 CACATGGATGGCTTGAATTTCTTCGCGAGCTCTGAGGCTGCTAAAGTTTCCCGCTGG 1837
Qy 935 TCTGTGTCCACTGCGCACTGCTTTTACGAGCGGCTACCTCACCGGAGGAGCAGCTCCG 994
Db 1838 CCATTTGGAGTATCCGGAAGATATATGGAATTTGGGAGAAATTTCCCAAGCAGCTTC 1897
Qy 995 CTTCAATTAATTTTACCTTGTCTCTTAAAGGACCTTCCGAGACCAAGATCTATT 1054
Db 1898 ATCCCTGGTCAACTACACTGACCAACACGACCCCTATGCTGAACCTCAGGTGTATT 1957
Qy 1055 CCCTGCTTTTGGGAGAACACAAACCAATCGCGAGGATTTGCCACCTTCTTTGAGAG 1114
Db 1958 CACGTTTTCGATGATGAGCCGGAATCAGCAATGCTTTGACATCTTCTTCCACG 2017
Qy 1115 CAGAGTTGGGTGGCTTGGCTAAGACTATC 1146
Db 2018 TCAGGATTTGACGACATGGCGAAAAAGTACC 2049

RESULT 11
AY262014 1783 bp DNA linear PLN 12-DEC-2003
LOCUS Claviceps purpurea dimethylallyltryptophan synthase (dmaW) gene,
DEFINITION dmaW-2 allele, partial sequence.
ACCESSION AY262014
VERSION AY262014.1 GI:32967583
KEYWORDS Claviceps purpurea (ergot fungus)
SOURCE Claviceps purpurea
ORGANISM Claviceps purpurea
Eukaryota; Fungi; Ascomycota; Pezizomycetina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
REFERENCE 1 (bases 1 to 1783)
AUTHORS Wang,J., Machado,C. and Scharld,C.L.
TITLE The determinant step in ergot alkaloid biosynthesis by a grass
endophyte
JOURNAL Fungal Genet. Biol. (2004) In press
REFERENCE 2 (bases 1 to 1783)
AUTHORS Wang,J., Machado,C. and Scharld,C.L.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2003) Plant Pathology, University of Kentucky,
201 Plant Science Building, Lexington, KY 40546-0312, USA
FEATURES
1. .1783
/organism="Claviceps purpurea"
/mol_type="genomic DNA"

/isolate="ATCC 20102"
/specific_host="Secale cereale"
/db_xref="ATCC:20102"
/db_xref="taxon:5111"
140. .>1609
/gene="dmaW"
/allele="2"
/note="dimethylallyltryptophan synthase; null allele due
to frameshift"
275. .281
/gene="dmaW"
/allele="2"
/note="insertion; probable transposon footprint"

ORIGIN
Query Match 13.9%; Score 193.6; DB 8; Length 1783;
Best Local Similarity 52.1%; Pred. No. 3.8e-46;
Matches 580; Conservative 0; Mismatches 46; Indels 40; Gaps 5;

Qy 67 TGAGTCAAGCCTGAACTCTCGAACAGGACACACAAATGTGTATAGCAGTC 126
Db 183 TGATCTGATTTTGAGTTCCCGCATGACAGCGACTATGTTGGCAGTACCGCGC 242
Qy 127 CGATGTTTSCACCATGATGGCGGGCGGCTATGATGTT-----CACGCAAGTAC 179
Db 243 CTATGTTTGGCGGATGCTTGACAATGCCGCTACACTGTACAGTGTCCACGACCAATAT 302
Qy 180 AGTTCTCTGTATCCACGTTGAGTGTATCATCCCGGCTTGGGTCCATACCCAGAAAAG 239
Db 303 CGGCATCTGAGCATTTTCAAGACACATTAATCCCTTTCTTGTGTGTATCCACAAAA 362
Qy 240 GGTGAGCCATGCTCACTGGAAGAGTCTCTACACGCTTCGGACTTCTTTTCGAGCTGAGC 299
Db 363 GACAAGGAAAG---ATGGCTCAGCATCTCCAGATCGGCTCTTCTTTTGGAACTGAGT 419
Qy 300 TTCAATTAATCAATCACTACTAGGTTTTCGATTCGAGCCCTCGGTTTCCCTGACGGGA 359
Db 420 TTGAATTTGATCCGATCTGTGTTTTCGATATATACAGAGCCCATCAATGAGATGACGGGG 479
Qy 360 ACGAAGGATGATCCATTCACACCCAGCAATCAGGCTGTTCTCCAGACCTCAAGGCC 419
Db 480 ACGGAGAAAGATCCATTCATACGTTGCCGATATAGGAAGTGTCCAAAGCTTCCCGCAG 539
Qy 420 ATGTTTCCAGGCTTTCGATGATGTTTCGATCAATTCATTAAGCAATTTGTTGTTTCG 479
Db 540 ATTCAAGCGGTTATCGACTTGGAGTGTGTTTACTTCAAGGATGAGTTCGCTTGGAC 599
Qy 480 GAGGAGAGGCTCGGACTCTGCTAGATCGATATTTAGATCCCGCTTTCAGAGCACAG 539
Db 600 GAATCGGATCGGCCATCTTCAAGATACCTGAGTGTGTCAGAGGACAGATAAAGACGAG 659
Qy 540 AACAACTGGCAGCCGATCTGGAGCATCTGCGGATATTGTTTGAAGACCTTACATCTAC 599
Db 660 AACAGTTAGCTTGGATCTCAA---AGAACCGAGTTCCGCTCAAAGTTTACTTCTAT 716
Qy 600 CCGGGATCAAGTCGATCGGACCGGACCCCAAAGAGAGACTCATGTTTGAAGC----- 655
Db 717 CGGCATCTCAATCGATCGGACCGGCAATCCACACACTTTTCTCATCTTTGACTCCGTG 776
Qy 656 -----AATCAAGGCTGCGACAAAGTTTGGCAAGTTTGGCAAGTTTGGCACTCCACTGGC 700
Db 777 TCAAGTTTTCGAGAGCATGACAGCATACAGCCCGCTTCCAGGCAATTTGCGACTAT 836
Qy 701 AATCTCGAGGAGTTTATAGTGTAGCGAGCACCCA-----CCCTCTCGGCCACTTT 752
Db 837 GTTTCGCGGCGAAATTAATTCGCGAGATAGACCAACAGGAGCCCTACATGCGGCTCTC 896
Qy 753 CTCTCATGAGTTTGGTCAAGCGCTCGAGTCCCGAATCAAGGTTCTACTGTATGGAAGC 812
Db 897 TTCTCATGCGATTTGATCGATCCCGCAAGTCTCGCGTCAAGATATACCTGCGAGGAGAG 956
Qy 813 CAGCTCGACCTGGCTCCATCGAAGTATTTGGACTCTCAAGCGGCGAGCAAGCATCCA 872

Db 957 ACGGTCTCAATGTCGCCGATGGAAGATCTGTGACCGCTGGCGCGCGACAAAGTCGACGCA 1016
 Qy 873 GAGACACTGGATGCTCGGATCGCTGAGGAGCTGTGCGAGCTAATGCCCGTCACGGAG 932
 Db 1017 TCCACCACTGGATGCTGAAATCTTCGCGAGCTCTGGAGCTGTAAAGTTCGCGCT 1076
 Qy 933 GGTCTGTGTCACATGCGCACTGCTTTTACGAGCGGCTACCTCACCGAGGAGAGCTC 992
 Db 1077 GGCACCTTGGATGATCGGAGGAGATATGGAATTTGGGAGAAATTCACACGAGCAGCTT 1136
 Qy 993 CCCTTCATTAATAATTTACCTGTCTCTTAAAGGCGCACTCCCGAACACACAGATCTAT 1052
 Db 1137 CCATCCTGTCTCACTACACTGACCAACAGACCCCATGCTGAACTTCAGGTGTAT 1196
 Qy 1053 TTCCCTGCTTTTGGCGAGACGACAAACCATCGCGAAGATGCGCACCTTTCTTTGAG 1112
 Db 1197 TTACCGGTTTTCGCGATGATGACGCGGAAATCAGCAATGCTTTGACCATCTTCTTCCAA 1256
 Qy 1113 AGCAGAGTTTGGGTGCTGCTAGAGCTATC 1146
 Db 1257 CGTCACGATTTGACGACATGGCGAAAAGTACC 1290

RESULT 12
 CLCDMAW 1894 bp DNA linear PLN 04-JUN-1998
 LOCUS Claviceps fusiformis dimethylallyl-trans-transferase (dmaW) gene,
 DEFINITION complete cds.
 ACCESSION L39640.1 GI:1005417
 VERSION 1
 KEYWORDS dimethylallyl-diphosphate; dimethylallyltransferase; dmaW gene; prenilyltransferase; tryptophan.
 SOURCE Claviceps fusiformis
 ORGANISM Claviceps fusiformis
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
 REFERENCE 1 (bases 1 to 1894)
 AUTHORS Tsai H.F., Wang H., Gebler J.C., Poulter C.D. and Schardl C.L.
 TITLE The Claviceps purpurea gene encoding dimethylallyltryptophan
 SYNTHASE, the committed step for ergot alkaloid biosynthesis
 JOURNAL Biochem. Biophys. Res. Commun. 216 (1), 119-125 (1995)
 MEDLINE 96067540
 PUBMED 7488077
 FEATURES
 Location/Qualifiers
 1..1894
 /organism="Claviceps fusiformis"
 /mol_type="genomic DNA"
 /db_xref="taxon:40602"
 /clone_lib="H-F Tsai's pMOcosX"
 /dev_stage="mycelium"
 /tissue_lib="ATCC 26245"
 161..1894
 /gene="dmaW"
 join(161..1381,1456..1577,1640..1894)
 /gene="dmaW"
 /product="dimethylallyl-trans-transferase"
 161..1381
 /gene="dmaW"
 /number=1
 /evidence=experimental
 join(195..1381,1456..1577,1640..1894)
 /gene="dmaW"
 /EC_number="2.5.1.1"
 /standard_name="DMAT synthase"
 /codon_start=1
 /product="dimethylallyl-trans-transferase"
 /protein_id="AAC1893.1"
 /db_xref="GI:1005418"
 /translation="MMTKAPATAVYDTLSLLFFNQEORLWVHIAIPVFAMLDITAG
 HNVHDIYRLHGIKGIHIFPLGYPAGQKHWPVSVLTRYGIFPELSLNCDSVRYTF
 EPTTEHTGDDSYNAFLEICTQKLVRIQPGIDMEWFSYFRNELVFNATARSIGRN
 DSVNQPIRTONKALDLKGRDFALKVLYPLHKLSIATGVSSHDLIENSVKLSQKHT

intron 1382..1455
 /gene="dmaW"
 /number=1
 /evidence=experimental
 exon 1456..1577
 /gene="dmaW"
 /number=2
 /evidence=experimental
 intron 1578..1639
 /gene="dmaW"
 /number=2
 /evidence=experimental
 exon 1640..1894
 /gene="dmaW"
 /number=3
 /evidence=experimental

ORIGIN

Query Match 13.1%; Score 182.6; DB 8; Length 1894;
 Best Local Similarity 51.5%; Pred. No. 7.6e-43;
 Matches 586; Conservative 0; Mismatches 494; Indels 57; Gaps 5;
 Qy 61 ACGTGTGAGTCAAGCCCTGAACCTCTCGAAGAGGACACACAAAATGTTGATATGAC 120
 Db 229 ACACCTGAGTCTGCTTTTTCATTTCCCAATCAAGAGACGCGACTTTGTTGGCAGTA 288
 Qy 121 CAGCTCCGATGTTGCCACCATGATGCGGGGGCGGCTATGATGTTTCACGACAGTACA 180
 Db 289 TCGGCCCATGTTTTCGGCCATGCTCGATCTGTGACACACAGTTTCACGACCAATATC 348
 Qy 181 AGTTCTCTGTATCCACCGTGAAGTCAATCCCGGGGCTTTGGGTCCATATCCAGAAAAG 240
 Db 349 GACACCTGGGATTTTCAAGAAGCACATCATCCCTTTCTCTGGGGGTCTATCCAGCGCAAG 408
 Qy 241 GTCAGCCCATGCACTGGAAGAGTCATCTCACGCTTCGAGCTTCCTTTTCGAGCTGAGCT 300
 Db 409 GAAAGACACA---CATGGCCCGATGTCCTCACAGATATGGCATAACCGTTTCAACTGAGCC 465
 Qy 301 TCAATTACTCCAAATCACTACTACGGTTTTCGATTTCGAGCCCGCTCGTTCCTCGACGGAA 360
 Db 466 TCAACTGCTGACTCGTTCGCTGCTGACATTCGAAACCCACACTGACACACTGGCA 525
 Qy 361 CGAAGGATGATCCATTCAACACCGAGCAATCAGGCGCTGTTTTCAGGAGCTCAAGGCCA 420
 Db 526 CAGGTGATGATTCATACACGCAATTTTCGATTTCTGGAATGTATCCAAAAGCTTTGTCGGA 585
 Qy 421 TGGTTCCAGGCTTGAACCTGGAATGTTTCGATCATTTCAAGAGCATTCGTTTCG 480
 Db 586 TCCAGCCGGGATCGACATGAGTGGTTTCAGTACTTTCAGGATGAGTTGTTCTGACG 645
 Qy 481 AGGAAGAGGCTCGGAGCTCTGCTAGATCGAGATATTGAGATCCCC---GTCTCAAGACAC 537
 Db 646 CGACGGAATCTGCACGCTTTGGACCTAACGACTCGGTGAACCAACCAACCATACGGACCC 705
 Qy 538 AGAACAACTGGGAGCGGATCTGGGCCATCTCGCGATATTGCTTGAACCTACATCT 597
 Db 706 AGAACAACTGGGCTTGGATCTCAAGGGCGATCG---TTTCGCGCTCAAGGCTCTATCTCT 762
 Qy 598 ACCCGCGGATCAAGTTCGATCGGACCGGACCCCAAGAGAGACTCATGTTTTCAC--- 653
 Db 763 ACCCTCATCTCAAGTCAATTTGCCACCGCGCTTTTCGTTCACACGACCTCATATTCAACTCGG 822
 Qy 654 -----GCAATCAAGGCTCCGACAGTTTGGCAAGTT 686
 Db 823 TGCGCAAGCTGTGCGAAGACACACTAGTATTGAGCCCTCTTCAACGTTTGTGGACT 882
 Qy 687 GCCACTCCTGCGCAATCTCTC-----GAGGAGTTTATAGCTGAGCGAG 729

Db 883 ATGTGCGCTCGGAAACGATCCGATTGCAATGCACGACGAGCAAGACGAGGATTCAG 942
 QY 730 CACCCACCTCTCGGCCACTTTCTCTCATGCGATTGCTCAAGCGCTCGAGTCCCGAA 789
 Db 943 CGAGCGCTTACGAGCGGCACTCTTGTGCTGACTTGTGCTGATCGTCCAAATCTCGCA 1002
 QY 790 TCAGAGTCTACTGTATGGAAGCCAGCTGACCTGSCCTCCATCGAAGTATTTGAGTTC 849
 Db 1003 TCAAGATATACTTTGTGGAGCAGACAGTCTCTTGTGACCGCAATGGAAGATCTGTGAGCT 1062
 QY 850 TCAACGGGGACGGAACGATCCAGAGACACTGATGCTGTGATGCGCTGAGGGAGCTGT 909
 Db 1063 TGGGGGACGAGAACCGACTCTTCCACCTGAACGGGCTTGACATGATGCGAGAGCTTT 1122
 QY 910 GCGAGCTATGCGCGTCAAGAGGCTGTGTGCTCACTGCGGACTGCTTTTACGAGCGCG 969
 Db 1123 GGCATCTTTCGAGATTCGCTCGGCTTCATGAAGTATCCAGATCTGACTTTGAAGCTGG 1182
 QY 970 GTACCTCACCGAGGAGCTCCCTTTCATATAAATTTTACCTTGTCTCTTAAAGCG 1029
 Db 1183 GCGAAGTCCCGAGAGAGCTTCGTCATGCTCCATTTATGCTTGCATCCGACCAAC 1242
 QY 1030 CACTTCCGAAACACAGATCTATTTCCCTGCTTTTGGGAGAACGACAAACCATCGCGG 1089
 Db 1243 CAATGCCGAGCCACAGGTGATTTTACAGTGTGTTGGCATGAGTGACGCGGAATCACCA 1302
 QY 1090 AAGGATGCGCACTTCTTTAGAGCAGAGGTTGGGCTGCTTGGCTAAGAGCTATC 1146
 Db 1303 ATGCTTGGCCACTTCTTACGCCGACACGATGATGAGATGGCAAAAAGTACC 1359

RESULT 13
 AR181515
 LOCUS AR181515 1598 bp DNA linear PAT 20-APR-2002
 DEFINITION Sequence 6 from patent US 6335188.
 ACCESSION AR181515
 VERSION AR181515.1 GI:20223729
 KEYWORDS
 SOURCE Unknown.
 ORGANISM
 Unclassified.
 1 (bases 1 to 1598)
 Schardi,C.L. and Wang,J.
 Endophyte ergot alkaloid synthetic compounds, compounds which
 encode therefor and related methods
 Patent: US 6335188-A 6 01-JAN-2002;
 JOURNAL Location/Qualifiers
 FEATURES
 source 1..1598
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN
 Query Match 12.7%; Score 176.6; DB 6; Length 1598;
 Best Local Similarity 50.7%; Pred. No. 4.7e-41;
 Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;
 QY 33 GCAGCAACACTGTGCGAAAGCCCTTCTACGTGCTGAGTCAAGCCCTGAACTCTCGAAC 92
 Db 34 GCAAGACACTCCACCAGGAAGTTTATCAAAACCTCAGCGAAACATTTGACTTTGCCAAC 93
 QY 93 AAGGACACACAAATGCTGATAGACAGCTCCGATGTTGCCACCATGATGGGGG 152
 Db 94 AATGACCAAGGCTATGGTGGCAGCAGCGCGCCCAATGTTCCAAAGATCTCCAACT 153
 QY 153 GCGGCTATGATGTTACGACACAGTCAAGTTCCTCTGTATCCACCGTGAAGTCAATC 212
 Db 154 GCTAACTATAGCATTTATGCTAGTATCAACATCTGAGCACTTATAAAGCCATATCA 213
 QY 213 CCGCGTGGTCCATACCCAGAAAGGCTGACCCCATGCACTGGAAGATCATCTACA 272
 Db 214 CTTTCTTGGTGTCTATCTTCAAGAAGTGG---CGAGCGGTGGCTTAAGCATTTCTAG 270
 QY 273 CGCTTCGAGCTTCTTTTCGAGCTGAGCTTCAATTACTTCCAAATCACTACTACGTTTGA 332

Db 271 AGATACGGAACCCGTTTGAGCTTAAGTCTTAATTTGCTCTGACTCCATAGTTCCGTATACA 330
 QY 333 TTCAGAGCCCTCGGTTCCCTGACGGGAAACGAGAGATGATCCATTCACACCCAGGCAATC 392
 Db 331 TACGAGCCTATTACGCGCGCAACTGTCGAGCCATCTGATCCGTTTCAACACTTTTCGCTATC 390
 QY 393 AGGCTCTTCTCCAGGACCTCAAGGCCATGTTTCCAGGGCTTGACCTGGAAATGTTCCGAT 452
 Db 391 TGGAGGCTCTTAAAAAAGCTTATAGATCCAGCCAGGCATAGACCTTCATGTTTTC 450
 QY 453 CATTTCACTAAGCAATTTGGTTCGAGAGAAAGAGCTCGGACTCTGTCTAGATCGAGAT 512
 Db 451 TACTTTAAACAGAGCTTACACTTGACGCAAAACAGTCCAGTCCAGTACCTCTCTCAAAAC 510
 QY 513 ATTGAGATCCCGCTCTTCAGACACAGAACTGCGAGCCGATCTGGAGCCATCTCGGC 572
 Db 511 TTGTGCAAGAACAGATCAAAACCTCAAAACAGTTCGCTTACCTTAA---AGGTGAC 567
 QY 573 GATATTGTTTGAAGACTTACATCTACCCGCGGATCAAGTCGATCGGACCGGACCCCA 632
 Db 568 AAGTTCGTACTCAAGACTTACATCTACCCGCAATTAAGTCCGTCGCAACTGTTAAATCG 627
 QY 633 AAAGAGAGACTCATGTTTGGAGCAAT---CHAGGCTGCGGACAGTTCGCAAAAGTTGCC 689
 Db 628 GTCCAGGAGCTTGTGTTTGGCTCCGCAAGCTAGCGAGAGCATTAAGATATCCGT 687
 QY 690 ACTCCACTGGCAATCTTCGAGGAGTTTAT--AGCTGAGCGAGACCCACCTCCTCG-- 745
 Db 688 CTTGCTTTGAATGCTAGAGACTATGTCCAGTCTCGCAATAAGTCCCTTACCACGGAT 747
 QY 746 -----CCACTTTCTCTATGCGATTTGGTCAAGCCGCTCC 779
 Db 748 GACAGTCACAATACTCCATTTATCTTCACGCTTCTCTTTCGCGACCTGGTGAAGTCTTACC 807
 QY 780 GAGTCCCGAATCAAGGCTACTGTATGAGCGCCAGCTCGACCTGGCTCCCTCATCGAAGGT 839
 Db 808 AAGTCTCGTCAAGATCTACCTCTGGAACGAATGTTCTGTTGCCAGCATGGAAGAT 867
 QY 840 ATTTGGACTCTCAACGGCGGACGGAACGATCCAGAGACACTGATGTTCTGGATGCGGTG 899
 Db 868 CTTTGGAGCTTGGCGCGTTCGAGAAGATCAGTCCACTATTGAGGGATTGGAGATGATC 927
 QY 900 AGGAGCTGTGGAGCTATTTCGCCGTACGAGAGGCTGTGTCTCACTGCGCAACTGCTTT 959
 Db 928 CGAGAACTTTGGGGTCTCTTAAACATGCTCTCTGTTGCGCGCTTACCTTGAGCCCTTAC 987
 QY 960 TAGGAGCGGCTACCTCACCGCAGAGCTCCCTTCAATTATAAATTTTACCTTTGTCT 1019
 Db 988 TTGCCCTCGGGCCATTTCCCAATGAGCACTTCGTCATGCGCAATTACCTTACAC 1047
 QY 1020 CTTAAAGCGCACTTCCGAAACCAAGATCTATTTCCTGCTTTTGGGCAAGACGACAAA 1079
 Db 1048 CATATGATCCGATACCAAGAACCGCAAGTGTACTTTACTGTGTTCCGATGAATGATG 1107
 QY 1080 ACCATCGGGAAGGATGGCCACCTTCTTTCGAGAGCAGAGGTTGGGTGGCTTGGCTAAG 1139
 Db 1108 GAGGTGACTAATGCACTCAGCAATTTCTTCATGAGGCATGAATGAGCGATATGCGAAGT 1167
 QY 1140 AGCTATCCAGC 1150
 Db 1168 AATATACAGC 1178

RESULT 14
 AY259839
 LOCUS Neotyphodium coenophialum 6055 bp DNA linear PLN 12-DEC-2003
 DEFINITION Neotyphodium coenophialum L-tryptophan dimethylallyl transferase (dmaw) gene, dmaw-2 allele, complete cds.
 ACCESSION AY259839
 VERSION AY259839.1 GI:32402651
 KEYWORDS
 SOURCE Neotyphodium coenophialum

ORGANISM	Neotyphodium coenophialum
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Neotyphodium.
AUTHORS	1. (bases 1 to 6055)
TITLE	Wang, J., Machado, C. and Schardl, C.L.
JOURNAL	The determinant step in ergot alkaloid biosynthesis by a grass endophyte
REFERENCE	Fungal Genet. Biol. (2004) In press
AUTHORS	2. (bases 1 to 6055)
TITLE	Machado, C., Wang, J. and Schardl, C.L.
JOURNAL	Direct Submission
FEATURES	Submitted (17-MAR-2003) Plant Pathology, University of Kentucky, 201 Plant Sciences Bldg., Lexington, KY 40546-0312, USA
source	Location/Qualifiers
1. .6055	/organism="Neotyphodium coenophialum"
	/mol_type="genomic DNA"
	/isolate="e19; ATCC90664"
	/specific_host="Lolium arundinaceum (synonym:Pestuca arundinacea)"
gene	/db_xref="ATCC:90664"
	/db_xref="taxon:5047"
	<2961. .>4447
	/gene="dmaW"
	/allele="dmaW-2"
mRNA	join(<2961. .4132,4199. .4320,4389. .>4447)
	/gene="dmaW"
	/allele="dmaW-2"
CDS	/product="L-tryptophan dimethylallyl transferase"
	join(2961. .4132,4199. .4320,4389. .4447)
	/gene="dmaW"
	/allele="dmaW-2"
	/EC_number="2.5.1.34"
	/notes="aromatic prenyltransferase;
	4-(gamma,gamma)-dimethylallyltryptophan synthase; DMAT synthase; 4-(3-methylbut-2-en-1-yl)-L-tryptophan synthase; dimethylallyl-pyrophosphate"
	/codon_start=1
	/product="L-tryptophan dimethylallyl transferase"
	/protein_id="AA881208.1"
	/db_xref="GI:32402652"
	/translation="MVLAKTLHQEVYQTLSETFDANRDLRWHSTAPMFQKILQTA NYGIAYQHLSTYKSHIIPFLGVPYTRSGRWLSILTRYGPFSLINLCSIVRYT YEPINAAQSHLDPFNFTALWEALKLIDSGPDLQWFSYFKOELTLDANESTYLSH ONLVKQIKTONKALDLKGDKEVLTVIYBELKSVATGSKVQELAVFGSVKLAQKH SIRPAPEMLEDVQSNKVPYTDSDNTEPLSRLSLCDLVSPFKSRVKIYLLERWVSL FAMEDLWTIGRRDQSTLEGLEIMRELWGLNMGSPGLRAYPEPYLPGLAIPNEQLPS MANYTLHNDPIPEQVYTFVGMNDMEVTNALTFRFMEHWSDMASKYKALRESFP HNYEALNVIHSYISFYSRNNRPYLSVYLSHSPETGWPVPVPEGLIAFDGCRDLTCYK "
ORIGIN	
Query Match	12.7%; Score 176.6; DB 8; Length 6055;
Best Local Similarity	50.7%; Pred. No. 5.3e-41;
Matches 583; Conservative	0; Mismatches 529; Indels 39; Gaps 5;
QY	33 GCAGCAACACTGCTGCGCAAGCCCTTCTACGTGCTGAGTCAAGCCCTGAACCTCTCGAAC 92
Db	2970 GCAAGCACTCCACAGGAGTTTATCAACCCCTCAGCGAAACATTTGACTTTGCCAAC 3029
QY	93 AAGGACACACAAATAGTGGTATAGACACAGCTCCGATGTTTGGCCACCATGATGGGGG 152
Db	3030 AATGACCAAGGCTATGGTGGCACAGCAGCGGGCCAAATGTTCCAAAGATATCTCAAACT 3089
QY	153 GCCGGGTATGATGTTCCAGCACAGTACAAAGTTCTCTGTATCCACCGTGAGTCAATC 212
Db	3090 GCTAACTATAGCATTTATGCTTGGTATCAACATCTGAGCATTTTATAAAGCCATATCAIT 3149
QY	213 CCGCGGTGGTCCATACCCAGAAAGGCTGAGCCCATGCACTGGAAGAGTCAATCTCA 272
Db	3150 CTTTTTCTTGGTGTCTATCTCATCAAGAGTGG---CGAGCGGTGGCTAAGCATTTCTTACG 3206

QY	273 CGCTTCGAGCTTCCTTTTCGAGCTGAGCTTCAATTACTCAAAATCACTACTACGCTTTCGA 332
Db	3207 AGATACGGAACCCCGTTTTCGAGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 3266
QY	333 TTCGAGCCCTCGGTTCCCTGACCGGGAAGAAAGATGATCCATTCACACCCAGGCAATC 392
Db	3267 TACGAGCCCTATTAACGCGCGCAACTGGGAGCCATCTGGATCCGTTTCAACACTTTCGCTATC 3326
QY	393 AGGCTGTTCCTCCAGGACCTCAAGGCCATGTTCACGGGCTTGACCTTGGAAATGTTTCGAT 452
Db	3327 TGGGAGGCTCTAAAAGAGCTTATAGATTTCCAGCCAGGCAATAGACCTTCAATGTTTTC 3386
QY	453 CATTTCACTAAAGCATTTGGTTCGTTTCGGAGGAAGAGCTCGGACTCTGCTAGATFCGAGAT 512
Db	3387 TACTTTAAACAAGAGCTTACACTTTGACGCAACGAGTCCAGTACCTGCACTCTCAAAAC 3446
QY	513 ATTGAGATCCCGCTTCACAGACACAGAACAACTGGCAGCCGATCTGGAGCCATCTGGC 572
Db	3447 TTGCTCAGGAACAGATCAAAACTCAAAACAGCTAGCTTACCTTAACTTAACTTAACTTAACT 3503
QY	573 GATATTGCTTTGAAGACTACATCTACCCGCGGATCAAGTCGATCGCGACCGGACCCCA 632
Db	3504 AAGTTGCTTCAAGACTACATCTACCCGGAATTTGAAGTCCGTCGCAACTGGTAAATCG 3563
QY	633 AAAGAGAGACTCATGTTTGACGCAAT---CAAGSCTCCGACAAAGTTTGGCAAGTTGCC 689
Db	3564 GTCCAGGAGCTTTGTTTGGCTCCGTCGCAAGCTAGCGAGAGCAATAGAGATATCCGT 3623
QY	690 ACTCCACTGGCAATCTCTCGAGAGTTTAT---AGCTGAGCGAGCACCCACCTCTCTCGG-- 745
Db	3624 CTGCTTTGAAATGCTAGAGACTATGTCAGTCTCGCAATAAGTCCCTACACCGAT 3683
QY	746 -----CCACTTCTCTCATCGGATTTTGGTCAAGCGCTCC 779
Db	3684 GACAGTCACAATACTCCATTATCTTCAAGCGCTTCTCTTTCGACCTTGTGAGTCTCTTACC 3743
QY	780 GAGTCCCGAATCAAGGCTTACTGTATGGAACGCCAGCTCGACCTGGCTCCATCGAAGGT 839
Db	3744 AAGTCTGCTCAGATCTACTCTCTGGAAGATGCTCTGTTGCGAGCGATGGAAGAT 3803
QY	840 ATTTGGACTCTCAACGGCGACGATCCAGACACTCGAGACTGATGCTTGGATGGCGCTG 899
Db	3804 CTTTGGACGCTTGGCGCGCTCGAGAAGATCAGTCCACTATTGAGGATTTGAGATGATC 3863
QY	900 AGGAGAGTGGCAGCTATTGCCCGTCACGAGGGTCTGTCTCACTGCCGAACTGCTTT 959
Db	3864 CGAAACTTTGGGGTCTCTTAAATGCTCTCTGTTTGGCGGCTACCCCTGAGCCCTTAC 3923
QY	960 TACGAGCGGGTACCTCACCGGAGGAGCTCCCTTCAATTATAAATTTTACTTGTCT 1919
Db	3924 TTGCCCTCGCGGCAATCCCAATGAGCAACTTCCGTCCTTCCCAATTAACCTTACAC 3983
QY	1020 CCTAAAAGCGACTTCCGGAACACAGATCTATTTCCCTGCTTTTGGGAGAGACGACAA 1079
Db	3984 CATAATGATCCGATACCAAGAACCCGCAAGTGTACTTTACTGTTCGSCATGAATGATG 4043
QY	1080 ACCATCGGGAAGGATTTGGCCACTTCTTTGAGAGCAGAGGTTGGGCTGGCTTAAG 1139
Db	4044 GAGTGACTAATGCACTCAGAAATCTTTCATGAGGCAATGAATGGAGCATATGGCAAGT 4103
QY	1140 AGCTATCCAGC 1150
Db	4104 AAATACAAAGC 4114
RESULT 15	
AY259837	AY259837 6640 bp DNA linear PLN 12-DEC-2003
LOCUS	Epichloe typhina x Neotyphodium lolii L-tryptophan dimethylallyl
DEFINITION	transferase (dnaW) gene, complete cds.
ACCESSION	AY259837
VERSION	AY259837.1
KEYWORDS	GI:32402647

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Epichloe typhina x Neotyphodium lolii
Epichloe typhina x Neotyphodium lolii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypotrachales; Clavicipitaceae; Epichloe.
1 (bases 1 to 6640)
Wang, J., Machado, C. and Schardl, C.L.
The determinant step in ergot alkaloid biosynthesis by a grass
endophyte
Fungal Genet. Biol. (2004) In press
2 (bases 1 to 6640)
Wang, J., Machado, C. and Schardl, C.L.
Direct Submission
Submitted (17-MAR-2003) Plant Pathology, University of Kentucky,
201 Plant Sciences Bldg., Lexington, KY 40546-0312, USA
Location/Qualifiers
1 6640
/organism="Epichloe typhina x Neotyphodium lolii"
/mol_type="genomic DNA"
/isolate="Lpl"
/db_xref="taxon:71426"
3111..>4595
/gene="dnaW"
join(<3111..4282,4347..4468,4537..>4595)
/gene="dnaW"
/product="L-tryptophan dimethylallyl transferase"
join(3111..4282,4347..4468,4537..4595)
/gene="dnaW"
/EC number="2.5.1.34"
/note="aromatic prenyltransferase;
4-(gamma,gamma)-dimethylallyl-tryptophan synthase; DWAT
synthase; 4-(3-methylbut-2-enyl)-L-tryptophan synthase;
dimethylallyl-pyrophosphate"
/codon_start=1
/product="L-tryptophan dimethylallyl transferase"
/protein_id="AAP81206.1"
/db_xref="GI:32402648"
/translation="MVLAKTHQEVYTLSETFDPAANDQRLWHSTAMPFOKILQTA
NYSIYQIQLSHYKSHLPFLGVPTFRGGERWLSLTRYGPFLSLNCSIVRYT
YEPINATGSHLPDFTFAEALKKLIDSPGIDQWFSYFKQLBLDANSTYHS
QNLVQIKITQKLDLKDQKLVLYTYPELKSATQSKVQELVFGSRKLAQKH
SIRPAPMEDYVQSRNKVPTDDSHNTPLSRLLSCDLVSPKSRVLYLERVSL
PAMEDLWTLGREDOSTEGLEMLRWGLNMSPLRAYPEPYLPGLAIPEQOLPS
MANYTLHNDPIPEPOVYTFVGMNDMEVNTALTKFMRHWSDMASKVKACLRSEFP
HNYEALNYLHSYISYSNNKPYLSVYLHSFETGWPVPEGLIAPDCRDLICYK
"

Query Match 12.7%; Score 176.6; DB 8; Length 6640;
Best Local Similarity 50.7%; Pred. No. 5.3e-41;
Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;

Qy 33 GCAGCAACACTGTCGCAAGCCCTTCTACGTGCTGAGTCAAGCCCTGACCTCTCGAAC 92
Db 3120 GCACAGACATCCACAGGAAGTTTATCAAAACCCCTCAGCGAAACATTTGACTTTGCCAAC 3179
Qy 93 AAGGACACACAAATGGTGTATAGCACAGCTCCGATGTTGCCACATGATGGCGGG 152
Db 3180 AATGACAGAGGCTATGGTGCCACAGACCGCGCCATGTTCCAAAGATCTCCAACT 3239
Qy 153 GCCGGCTATGATGTTACGACACAGTACAAGTTCCTGTGATCCACCGTGAAGTCAATC 212
Db 3240 GCTAACTATAGCATTTATGCTCAGTATCAACATCTGAGCATTTATAAAGCCATATCAAT 3299
Qy 213 CCGCGCTGGTCCATACCCAGAAAGGTTCAGCCCATGCTGGAAGATCATCTACA 272
Db 3300 CTTTTCTTGGTGTCTATCCGACAAAGTGG---CGAGCGGTGGCTAAGCATCTTACG 3356
Qy 273 CGCTTCGAGACTTCCTTTTCGAGCTGAGCTTCAATTAATCTCAAAATCACTACTACGGTTTGA 332
Db 3357 AGATACGGAACCCGGTTGAGCTAAGTCTTAATTGCTCTGACTCCATAGTTCGGTATACA 3416
Qy 333 TTCAGGCCCTCGGTTTCCTGACCGGGAACGAGAGATGATCGATTCAACACCCAGGCAATC 392

Db 3417 TACGAGCCTATTACGCCCAACTGCGAGCCATCTGGATCCGTTCAACACTTTGCTATC 3476
Qy 393 AGGCTCTGTTCCAGGACCTCAAGGCCATGGTTCAGGGGTTGACCTGGAAATGTTCCGAT 452
Db 3477 TGGGAGGCTCTAAAAAGCTTATAGATCCAGCCAGCATAGACCTTCAATGTTTCC 3536
Qy 453 CATTTCTACTAAGCATTTGGTGGTTTCGAGAGAGAGCTCGGACTCTGCTAGATCGAGAT 512
Db 3537 TACTTTAAACAAGAGCTTACACTTTGAGCGAAAACGAGTCCACGTACCTGCACTCTCAAAAC 3596
Qy 513 ATTGAGATCCCGCTCTCAAGACACAGAAACATGCGAGCCGATCTGGAGCCATCTGGC 572
Db 3597 TTGCTAAGGACAGATCAAAACTCAAAACAGAGTAGCTAGACCTTAA---AGGTGAC 3653
Qy 573 GATATTGCTTTGAAGACCTACATCTACCCGGGATCAAGTCGATCGGACGCGGACCCCA 632
Db 3654 AAGTTGCTGTTCAAGACCTACATCTACCCGAAATGAAGTCCGTCGCAACTGGTAAATCG 3713
Qy 633 AAAGAGACTCATGTTTTCAGCGAAT---CAAGGCTCCGACAAATTTGGCAAGTTGCC 689
Db 3714 GTCCAGGAGCTTGTGTTTGGCTCCGTCGCAAGCTAGCGCAAGATGAAGATTCGAT 3773
Qy 690 ACTCCACTGGCAATCTCGAGGAGTTTAT--AGCTGAGCGAGCACCCACCTCTCCGG-- 745
Db 3774 CTTGCCITTTGAATGCTAGAGACTATGTCCAGTCTCGCAATAAAGTCCCTTACCACGAT 3833
Qy 746 -----CCACTTTCTCTCATCGGATTTGGTCAAGCCGCTCC 779
Db 3834 GACAGTCACAATACTCCCAATATCTTCACGCTTCTCTCTTCGACCTGGTGAAGTCTTACC 3893
Qy 780 GAGTCCCGAAATCAAGGTCTACTGTATGGAACGCCAGCTCGACCTGGCTCCATCGAAGGT 839
Db 3894 AAGTCTGTGTCAGATCTACTCTCTGGAACGAATGGTCTGTTGCCAGCGATGGAAGAT 3953
Qy 840 ATTGGAATCTCAACGGCGACGGAACGATCCAGAGACACTGGATGGTCTGGATGCGGCTG 899
Db 3954 CTTTGGACGCTTGGCGCGCTCGAGAGATCAGTCCACTATTGAGGGATTTGAGATGATC 4013
Qy 900 AGGAGCTGTGGCAGCTATTGCCCGTCAAGGAGGTCTGTCCACTGCCGAACCTGCTTT 959
Db 4014 CGAGAACTTTGGGGTCTCCTTAAATGCTCTCCTGTTTGGCGGCTACCTCGAGCCTTAC 4073
Qy 960 TACGAGCCGGGTACCTCAACCGCAGAGCAGCTCCCTTCAATATAAAATTTTACCTTGCT 1019
Db 4074 TTGCCCCCTCGGGCCATTTCCCAATGAGCACTTCGCTCCATGGGCAATTACACCTTACAC 4133
Qy 1020 CCTAAAGCGACATTCGCCGAACACAGATCTATTTCCTGCTTTTGGSCAGAACACAA 1079
Db 4134 CATAATGATCCGATACCAAGACCCGCAAGTGTACTTGTGTTGGCATGAATGATATG 4193
Qy 1080 ACCATCGCGGAGGATTGGCCACCTTCTTTGAGAGCAGAGGTGGGGTGGCTTGGCTAAG 1139
Db 4194 GAGGTGACTAATGCACTACGNAATTCCTCATGAGGCATGAATGGAGCGATATGCAAGT 4253
Qy 1140 AGCTATCCAGC 1150
Db 4254 AAATACAAAGC 4264

Search completed: July 29, 2004, 00:10:31
Job time : 5592 secs

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 21:03:30 ; Search time 617 Seconds

(without alignments)
9591.151 Million cell updates/sec

Title: US-10-099-704-1

Perfect score: 1393

Sequence: 1 ccgaagctgacgaatggag.....cgaaaaaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1393	100.0	1393	3	AA51712 A. oryzae
2	1052	75.5	1053	3	AAFL14653 Aspergill
3	582.4	41.8	600	7	ABZ53042 Aspergill
4	205.2	14.7	1347	6	ABK15520 DNA encod
5	205.2	14.7	1353	6	ABK15521 DNA encod
6	176.6	12.7	1598	6	ABK15523 DNA encod
7	175.6	12.6	1676	3	AAFL1858 Aspergill
8	173.4	12.4	1908	6	ABK15522 DNA encod
9	98.6	7.1	657	3	AAFL12651 Aspergill
10	43.6	3.1	2000	7	ADA71938 Rice gene
11	38.8	2.8	2000	7	ADA71938 Rice gene
12	35.6	2.6	1258	7	ACA31530 Prokaryot
13	34.8	2.5	484	5	ABV14861 Human pro
14	34.8	2.5	3237	4	AA506331 DNA encod
15	34.8	2.5	4254	7	ACA42624 Prokaryot
16	34.6	2.5	4380	5	ABAL14731 Human ner
17	34.4	2.5	1065	7	ABV72510 Nucleotid
18	34	2.4	14041	4	AAH48024 Internal
19	34	2.4	349980	6	ABQ81846 Bifidobac
20	33.8	2.4	4060	6	ABQ81281 Arabidops
21	33.8	2.4	58557	3	AA558471 Nucleotid
22	33.6	2.4	339	2	AAQ11993 Transform
23	33.6	2.4	1071	3	AA424403 Arabidops

24	33.6	2.4	1303	2	AAQ09317 Monkey tr
25	33.6	2.4	1559	2	AAQ20289 Sequence
26	33.6	2.4	1560	1	AAAN81084 Coding se
27	33.6	2.4	1560	2	AAQ03508 Simian Tr
28	33.6	2.4	1561	2	AAQ04908 Sequence
29	33.6	2.4	1561	2	AAQ03268 Simian tr
30	33.6	2.4	1561	2	AAQ29177 TGF-beta
31	33.6	2.4	1571	2	AAQ03269 Human tra
32	33.4	2.4	489	6	ABL99528 Target ca
33	33.4	2.4	511	8	ACH17350 Human adu
34	33.4	2.4	998	9	ADD35866 Human hep
35	33.4	2.4	1326	6	AAD22696 Porcine t
36	33.4	2.4	11477	4	AAK68713 Human imm
37	33.4	2.4	11477	4	AAK71247 Human imm
38	33.4	2.4	11482	4	AAK68712 Human imm
39	33.4	2.4	11482	4	AAK71246 Human imm
40	33.4	2.4	96649	7	ACC45150 Human NAC
41	33.2	2.4	341	4	AAH72072 Human cer
42	33.2	2.4	350	4	AAH69197 Human cer
43	33.2	2.4	350	4	AAH71002 Human cer
44	33.2	2.4	3176	4	ABL05635 Drosophil
45	33.2	2.4	5415	4	ABL05634 Drosophil

ALIGNMENTS

RESULT 1

AA51712

ID AAA51712 standard; cDNA; 1393 BP.

XX

AC AAA51712;

XX

DT 31-OCT-2000 (first entry)

XX

DE A. oryzae dimethylallyl-cycloacetyl-L-tryptophan synthase cDNA.

XX

KW Dimethylallyl-cycloacetyl-L-tryptophan synthase; DCAT-S; beta-CPA;

KW toxin production; elimination; heterologous protein production;

XX filamentous fungi; ss.

XX Aspergillus oryzae.

XX

FH Key Location/Qualifiers

FT CDS 15..1328

FT /*tag= a

FT /product= "dimethylallyl-cycloacetyl-L-

FT tryptophan_synthase"

XX

PN WO200039322-A1.

XX

PD 06-JUL-2000.

XX

PF 22-DEC-1999; 99WO-DK000726.

XX

PR 23-DEC-1998; 98DK-00001726.

XX

PR 27-MAY-1999; 99DK-00000745.

XX

PA (NOVO) NOVO NORDISK AS.

XX

PI Christensen BE, Mollgaard H, Kaasgaard S, Lehnbeck J;

XX

DR WPI; 2000-452411/39.

XX

DR P-ESDB; AA96961.

XX

PT Producing a polypeptide of interest such as a hormone or enzyme,

PT comprising cultivating a mutant of a parent Aspergillus cell which

PT produces less of at least one toxin of interest compared to the parent

PT cell under the same conditions.

XX

PS Claim 3i; Page 59-61; 66pp; English.

XX

XX The A. oryzae dimethylallyl-cycloacetyl-L-tryptophan synthase (DCAT-

CC

CC S) is involved in the synthesis of beta-CPA, from cyclo-acetoacetyl-L-
 CC tryptophan and dimethylallylpyrophosphate, by its homology to a
 CC dimethylallyltryptophan synthase (DMAT-S) from *Claviceps purpurea*.
 CC *Aspergillus* host cells having a modification in the DCAT-S gene, leading
 CC to reduced or eliminated toxin production, are useful for expression of
 CC heterologous polypeptides of interest. Other toxins which may be reduced
 CC or eliminated comprise kojic acid, 3-nitropropionic acid, emodin, etc.
 CC The DCAT-S gene can be used to identify and disrupt similar genes in
 CC other filamentous fungal host strains such as *Trichoderma*, *Penicillium*
 CC and *Fusarium*
 XX
 SQ Sequence 1393 BP; 345 A; 387 C; 344 G; 317 T; 0 U; 0 Other;

Query Match 100.0%; Score 1393; DB 3; Length 1393;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGACGACACATCTGTCGCAAGCCCTTCT 60
 DB 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAAGACGACACATCTGTCGCAAGCCCTTCT 60
 QY 61 ACCTGTCTGAGTCAAGCCCTGAACCTCTCGAAACAAAGGACACACAAATGGTATAGCA 120
 DB 61 ACCTGTCTGAGTCAAGCCCTGAACCTCTCGAAACAAAGGACACACAAATGGTATAGCA 120
 QY 121 CAGCTCGATGTTGGCAACATGATGGCGGGCGCGCTATGATGTTACGGACACAGTACA 180
 DB 121 CAGCTCGATGTTGGCAACATGATGGCGGGCGCGCTATGATGTTACGGACACAGTACA 180
 QY 181 AGTTCTCTGATCCACCGTGAGTGCATCATCCCGCGCTGGGTCCATACCCAGAAAGG 240
 DB 181 AGTTCTCTGATCCACCGTGAGTGCATCATCCCGCGCTGGGTCCATACCCAGAAAGG 240
 QY 241 GTGAGCCCATGCACTGGAAGAGTCATCTCACAGCTTCGGAATTCCTTTGAGCTGAGCT 300
 DB 241 GTGAGCCCATGCACTGGAAGAGTCATCTCACAGCTTCGGAATTCCTTTGAGCTGAGCT 300
 QY 301 TCATTTACTCCAAATCACTACTACGTTTGCATTCGAGCCCTCGGTTCCCTGACGGGAA 360
 DB 301 TCATTTACTCCAAATCACTACTACGTTTGCATTCGAGCCCTCGGTTCCCTGACGGGAA 360
 QY 361 CGAAGGATGATCCATTCACACCCAGGCAATCAGGCTCTGTTCTCCAGGACCTCAAGGCA 420
 DB 361 CGAAGGATGATCCATTCACACCCAGGCAATCAGGCTCTGTTCTCCAGGACCTCAAGGCA 420
 QY 421 TGGTTCCAGGCTTGACCTGGAAATGTTGATCATTTTCACTAAGCATTTGCTTTCCG 480
 DB 421 TGGTTCCAGGCTTGACCTGGAAATGTTGATCATTTTCACTAAGCATTTGCTTTCCG 480
 QY 481 AGAAGAGGCTCGGATCTGCTAGATCGAGATATTCAGATCCCGCTTTCAGAGACACAGA 540
 DB 481 AGAAGAGGCTCGGATCTGCTAGATCGAGATATTCAGATCCCGCTTTCAGAGACACAGA 540
 QY 541 ACAAATCTGGCAGCGGATCTGGAGCCATCTGGCGATATTTCTTTGAAGACCTACATCTACC 600
 DB 541 ACAAATCTGGCAGCGGATCTGGAGCCATCTGGCGATATTTCTTTGAAGACCTACATCTACC 600
 QY 601 CGCGGATCAAGTCGATCGGACCGGACCCCAAGAGAGACATCATGTTTTCGCAATCA 660
 DB 601 CGCGGATCAAGTCGATCGGACCGGACCCCAAGAGAGACATCATGTTTTCGCAATCA 660
 QY 661 AGGCTGCCGCAAGTTTGGCAAGTTGGCACTCCCACTGGCAATTCCTTCGAGGAGTTTATAG 720
 DB 661 AGGCTGCCGCAAGTTTGGCAAGTTGGCACTCCCACTGGCAATTCCTTCGAGGAGTTTATAG 720
 QY 721 CTGAGCGAGACCCACCTCTCGGCACTTTCTCTCATGCGATTTGGTCAAGCGCTCG 780
 DB 721 CTGAGCGAGACCCACCTCTCGGCACTTTCTCTCATGCGATTTGGTCAAGCGCTCG 780
 QY 781 AGTCCCGAATCAAGGCTCTACTGATGGAACCGCAGCTCGACCTGGCTCCATCGAAGGTA 840
 DB 781 AGTCCCGAATCAAGGCTCTACTGATGGAACCGCAGCTCGACCTGGCTCCATCGAAGGTA 840

QY 841 TTTGGACTCTCAACGGCGGACGGAACGATCCAGAGACACTGGATGCTTGGATGCGCTGA 900
 DB 841 TTTGGACTCTCAACGGCGGACGGAACGATCCAGAGACACTGGATGCTTGGATGCGCTGA 900
 QY 901 GGGAGCTCTGGAGCTATTCCCGCTCACGAGGCTGTGTCCACTCCGAACTGCTTT 960
 DB 901 GGGAGCTCTGGAGCTATTCCCGCTCACGAGGCTGTGTGTCCACTCCGAACTGCTTT 960
 QY 961 ACAGCCGGGTACCTCACCAGCAGGACGAGCTCCCTTCATATAAAATTTTACCTTGTCTC 1020
 DB 961 ACAGCCGGGTACCTCACCAGCAGGACGAGCTCCCTTCATATAAAATTTTACCTTGTCTC 1020
 QY 1021 CTAAGAGCGGACTTCCGAAACGACAGATCTATTCCCTGCTTTTGGGACGACGACAAA 1080
 DB 1021 CTAAGAGCGGACTTCCGAAACGACAGATCTATTTCCTGCTTTTGGGACGACGACAAA 1080
 QY 1081 CCATCGCGGAAGGATTGGCCACCTCTTTGAGAGCAGAGGTTGGGCTTGGCTTAAGA 1140
 DB 1081 CCATCGCGGAAGGATTGGCCACCTCTTTGAGAGCAGAGGTTGGGCTTGGCTTAAGA 1140
 QY 1141 GCTATCCAGCGGATTTGGCATCTCTATATCCCGATGTGACCTGCGAATCACC 1200
 DB 1141 GCTATCCAGCGGATTTGGCATCTCTATATCCCGATGTGACCTGCGAATCACC 1200
 QY 1201 TGCAGGCGTGATCTCTCTTACAAAGGGGAAAAACCGTACATGAGTGTGACTCTCC 1260
 DB 1201 TGCAGGCGTGATCTCTCTTACAAAGGGGAAAAACCGTACATGAGTGTGACTCTCC 1260
 QY 1261 ATACCTTCGAAGGCTTCAGTGTCTGCTGCCAGAGGTTGCTATGTGACGATGGCCACA 1320
 DB 1261 ATACCTTCGAAGGCTTCAGTGTCTGCTGCCAGAGGTTGCTATGTGACGATGGCCACA 1320
 QY 1321 ATCCTTAGGACTAGTTTATCCCTTCATTCATGCATCCGTTGAATGTTGTCGAAAAA 1380
 DB 1321 ATCCTTAGGACTAGTTTATCCCTTCATTCATGCATCCGTTGAATGTTGTCGAAAAA 1380
 QY 1381 AAAAAAAAAAAAAA 1393
 DB 1381 AAAAAAAAAAAAAA 1393

RESULT 2

AAFL4653
 ID AAFL4653 standard; cDNA; 1053 BP.
 XX
 AC AAFL4653;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE *Aspergillus oryzae* EST SEQ ID NO:7176.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;
 KW *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS *Aspergillus oryzae*.
 XX
 PN WO200056762-A2.
 XX
 XX 28-SEP-2000.
 PD
 XX 22-MAR-2000; 2000WO-US007781.
 PF
 XX 22-MAR-1999; 99US-00273623.
 PR
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX

QY	541	ACAAACTGGCAGCGCATCTCGAGCCATCTCTGGCGATATTGTCTTGAAGACCTPACTCTACC	600
Db	541	ACAAACTGGCAGCGCATCTCGAGCCATCTCTGGCGATATTGTCTTGAAGACCTPACTCTACC	600
QY	601	CGCGGATCAAGTCGATCGCGACCGGACCCCAAAAGAGAGACTCATGTTTGACGCAATCA	660
Db	601	CGCGGATCAAGTCGATCGCGACCGGACCCCAAAAGAGAGACTCATGTTTGACGCAATCA	660
QY	661	AGGCTGCCGACAAAGTTTGGCAAAAGTTGCCACTCCACTGGCAATCCTCGAGGAGTTTATAG	720
Db	661	AGGCTGCCGACAAAGTTTGGCAAAAGTTGCCACTCCACTGGCAATCCTCGAGGAGTTTATAG	720
QY	721	CTGAGCGGACACCCACCCTCTCGGCCACTTCTCTCATCGGATTTGGTCAAGCCGCTCG	780
Db	721	CTGAGCGGACACCCACCCTCTCGGCCACTTCTCTCATCGGATTTGGTCAAGCCGCTCG	780
QY	781	AGTCCCGAATCAAGTCTACTGTATGGAACGCGCAGCTCGACCTGGCCCTCCATCGAAGGTA	840
Db	781	AGTCCCGAATCAAGTCTACTGTATGGAACGCGCAGCTCGACCTGGCCCTCCATCGAAGGTA	840
QY	841	TTTGGACTCTCAACGGCGAGCGGACGATCCAGAGACACTGGATGGTCTGATCGGCTGA	900
Db	841	TTTGGACTCTCAACGGCGAGCGGACGATCCAGAGACACTGGATGGTCTGATCGGCTGA	900
QY	901	GGGAGCTGTGGCAGCTATTGCCGCTCACGGAGGGTCTGTGTCCACTGCCGAACCTGCTTTT	960
Db	901	GGGAGCTGTGGCAGCTATTGCCGCTCACGGAGGGTCTGTGTCCACTGCCGAACCTGCTTTT	960
QY	961	ACGAGCGGGTACTCTCACCGGAGGACAGCTCCCTTCATTATAAAATTTTACCTTGCTTC	1020
Db	961	ACGAGCGGGTACTCTCACCGGAGGACAGCTCCCTTCATTATAAAATTTTACCTTGCTTC	1020
QY	1021	CTAAAGCGGCACCTTCCGGAACCCAGATCTATT	1053
Db	1021	CTAAAGCGGCACCTTCCGGAACCCAGATCTATT	1053
RESULT 3			
ABZ53042			
ID	ABZ53042	standard; cdna; 600 BP.	
AC	ABZ53042;		
XX			
DT	28-MAR-2003	(first entry)	
DE	Aspergillus oryzae	polynucleotide SEQ ID NO 2155.	
XX			
XX	Aspergillus oryzae;	fermentation; fungus; industrial; EST;	
KW		expressed sequence tag; gene; ss.	
OS	Aspergillus oryzae.		
XX			
PN	WO200279476-A1.		
XX			
PD	10-OCT-2002.		
XX			
PF	22-MAR-2002;	2002WO-1B000890.	
XX			
PR	30-MAR-2001;	2001JP-00098371.	
XX			
XX	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.		
PA	(NARE-) NAT RES INST BREWING.		
PA	(NORQ-) NAT FOOD RES INST MIN AGRIC.		
XX			
PI	Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;		
PI	Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;		
XX			
DR	WPI; 2003-046817/04.		
XX			
PT	Detection of expression of specific Aspergillus genes for monitoring the		
PT	fermentation and growth conditions of the fungus, using DNA probes.		
XX			

PS Claim 1; SEQ ID NO 2155; 48pp + Sequence Listing; Japanese.
 XX The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (AB250888-AB256893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridizing
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus,
 CC especially of *Aspergillus oryzae* which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 600 BP; 144 A; 176 C; 146 G; 134 T; 0 U; 0 Other;

Query Match 41.8%; Score 582.4; DB 7; Length 600;
 Best Local Similarity 99.8%; Pred. No. 2.9e-175;
 Matches 583; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 286 CTTTCGAGCTGACCTCAATTAATCAATCACTACTAGCTTGCATTCGAGCCCTCG 345
 DB 17 CTTTCGAGCTGACCTCAATTAATCAATCACTACTAGCTTGCATTCGAGCCCTCG 76
 QY 346 GTTCCCTGACGGGAACGAGGATGATCCATTAACACCCAGGCAATCAAGGCTGTTCTCC 405
 DB 77 GTTCCCTGACGGGAACGAGGATGATCCATTAACACCCAGGCAATCAAGGCTGTTCTCC 136
 QY 406 AGACCTCAAGGCGATGTTTCCAGGCTTGACCTCGAATGTTTGCATCAATTCACCTAAG 465
 DB 137 AGACCTCAAGGCGATGTTTCCAGGCTTGACCTCGAATGTTTGCATCAATTCACCTAAG 196
 QY 466 CATTGCTGTTTCGGAGGAGAGGCTCGGACTCTCTAGATCGAGATTTGAGATCCCGG 525
 DB 197 CATTGCTGTTTCGGAGGAGAGGCTCGGACTCTCTAGATCGAGATTTGAGATCCCGG 256
 QY 526 TCTTCAAGACACAGAACAACTGGCAGCCGATTCGAGCCATTCGCGATATTGTTCTGA 585
 DB 257 TCTTCAAGACACAGAACAACTGGCAGCCGATTCGAGCCATTCGCGATATTGTTCTGA 316
 QY 586 AGACCTACATCTACCCGGATCAAGTCGATCGGACCGGACCCCAAGAGAGACTCA 645
 DB 317 AGACCTACATCTACCCGGATCAAGTCGATCGGACCGGACCCCAAGAGAGACTCA 376
 QY 646 TGTTCGACGCAATCAAGGCTGCGACAAAGTTGGCAAAAGTTGCCATTCACCTGGCAATCC 705
 DB 377 TGTTCGACGCAATCAAGGCTGCGACAAAGTTGGCAAAAGTTGCCATTCACCTGGCAATCC 436
 QY 706 TCAGAGATTATAGTCGACGAGCACCACCCCTCTCGGCACTTCTCATCGATT 765
 DB 437 TCAGAGATTATAGTCGACGAGCACCACCCCTCTCGGCACTTCTCATCGATT 496
 QY 766 TGTTCGACGCAATCAAGGCTGCGACAAAGTTGGCAAAAGTTGCCATTCACCTGGCAATCC 825
 DB 497 TGTTCGACGCAATCAAGGCTGCGACAAAGTTGGCAAAAGTTGCCATTCACCTGGCAATCC 556
 QY 826 CTTCCATCGAAGTATTGAGCTCTCAACGGGCGAGCAACGAT 869
 DB 557 CTTCCATCGAAGTATTGAGCTCTCAACGGGCGAGCAACCAT 600

RESULT 4
 ABK15520
 ID ABK15520 standard; DNA; 1347 BP.
 XX
 AC ABK15520;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE DNA encoding dimethylallyltryptophan synthase (DmaW) version #1.
 XX

KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansiolepis; Echinodopsis; Atkinsonella; Myriogenospora;
 XX Neotyphodium; Parepichloe; knockout; antisense technology; gene; ds.
 OS Neotyphodium coenophialum.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1347
 FT /*tag= a
 FT /product= "dmaW"
 FT /note= "Dimethylallyltryptophan synthase"
 XX
 PN US6335188-B1.
 XX
 PD 01-JAN-2002.
 XX
 PF 03-MAR-2000; 2000US-00518657.
 XX
 PR 22-MAR-1999; 99US-0125490P.
 XX
 PA (KENT) UNIV KENTUCKY RES FOUND.
 XX
 PI Schardl CL, Wang J;
 XX
 DR WPI; 2002-163205/21.
 XX
 DR P-PSDB; AAU76412.
 XX
 PT New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 PT molecule) from fungi that are symbionts of commercially important
 PT grasses, useful to engineer ergot alkaloid-deficient symbionts.
 XX
 PS Claim 1; Col 9-12; 16pp; English.
 XX
 CC The invention describes an isolated dimethylallyltryptophan synthase
 CC (dmaW) nucleic acid (I). (I) is a dmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses
 CC and is useful for expressing dmaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a dmaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack dmaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of dmaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansiolepis,
 CC Echinodopsis, Atkinsonella, Myriogenospora, Neotyphodium, and
 CC Parepichloe, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This sequence encodes a
 CC dimethylallyltryptophan synthase (dmaW) described in the method of the
 CC invention
 XX
 SQ Sequence 1347 BP; 363 A; 345 C; 288 G; 351 T; 0 U; 0 Other;

Query Match 14.7%; Score 205.2; DB 6; Length 1347;
 Best Local Similarity 50.8%; Pred. No. 2.4e-54;
 Matches 647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;
 QY 33 GCAGCACACTGCTGCAAGCCCTTCTAGCTGCTGAGTCAGCCCTCGACCTCTCGAAC 92
 DB 10 GCAAGACACTCCACAGGAAGTTTATCATACCTTAGCGAAACATTTGACTTGGCAAC 69
 QY 93 AAGGACCAACAAAAATGGTGGTATAGCACAGCTCCGATGTTTCCACCATGATGGCGGG 152
 DB 70 AATGACCAGAGGCTATGGTGGCACAGCAGCGGCCCAATGTTGAAAAGATGCTCCAACT 129
 QY 153 GCGGCTATGATGTTTACGACAGTACAGTTCCTCTGTATCCACCGTAGGTCATCATC 212

Db 130 GCTAACTATAGCATTTGATGCTCAGTATCGACATCTGGGCAATTTATAAGAGCCATGTCATT 189
Qy 213 CGGGCGTTGGTTCATATACCCAGAAAGGTCAGCCATGCTGCACTGGAAGATGATCTACA 272
Db 190 CTTTTCCTTTGGTGTCTATCCTCAAGAAGTGG---CGAGCGGTGGCTAAGCATTTCTTACG 246
Qy 273 CGCTTCGGACTTCCTTTTCAGCTGAGCTTCAATTAATCTCAAAATCACTACTACGGTTTGA 332
Db 247 AGATACGGAAACCCGTTTGGCTAAGTCTTAATGCTCGACTCGTACTGCGTATATACA 306
Qy 333 TTGAGCCGCTCGTTTCCTGACGGGAACGAAGGATGATTCATTTCAACACCCAGGCAATC 392
Db 307 TAGGAGCCTATTAAACCCGCACTGCGAGTCTATCTGGATCCGTTTAAACATTTTCGCTATC 366
Qy 393 AGGCGTGTCTCAGAGACTCAAGGCCATGGTTCCAGGCTTTCAGGCTTTCAGTGAATGGTTGAT 452
Db 367 TGGAGGCCCTGAAGAAGCATATTGAGTCCGAGCCAGGATAGACTTGAATGGTTTCT 426
Qy 453 CATTTCACTAAAGCATTTGGTTTCGGAGGAGGCTCGGACTCTGCTAGATCGAGAT 512
Db 427 TACTTTAAACAAGAGCTTACACTTTGACGCAACGAGTCCACGTACTGCACTCGCAAAAC 486
Qy 513 ATTGAGATCCCGTCTTCAAGACAGACAACTGGCAGCCGATCTGGAGCCATCTGGC 572
Db 487 TTGGTTAAGAAACAGATCAAACTCAAAACAAGCTCGCTTTGGACCTTAA---AGGTGAC 543
Qy 573 GATATTGTCTTGAAGACCTACATCTACCCGCGATCAAGTCGATCGCGACCGGAGCCCA 632
Db 544 AAGTTCGTACTGAAGACCTACATCTACCCGGAATTTGAAGTCGCTCGCACTGCTAAATCG 603
Qy 633 AAAGAGAGACTGATTTTGAAGCAAT---CAAGGCTGCGCAAGTTCGGCAAGTTGCC 689
Db 604 GTCCAGGAGCTCGTGTGTGGCTCCGTCGCAAGTACGCGCAAGACCAAGAGTATTCGCT 663
Qy 690 ACTCCACTGCAATCTCTCAGGAGTTTAT--AGCTGAGGAGCAGCACCCCTCTCGG-- 745
Db 664 CCGTCTTTGAATGCTAGAAGACTATGTCAGTCTCGCAATAAATCTCTACACCGAT 723
Qy 746 -----CCACTTCTCTCATGCGATTTGCTCAAGCGTCC 779
Db 724 GACAGTCACAATACTCTATTATCTTCAAGCCTTCTCTTGGACCTGATAGTCTTACC 783
Qy 780 GAGTCCGGAATCAGGCTTACTGTATGGAAGCCAGCTCGACTGGCTCCATCGAAGT 839
Db 784 AAGTCTCGTCAAGATCTACCTCTCGGAACGAATGGTCTGTTGCGCAGGATGAAGAT 843
Qy 840 ATTGTGACTCTCAACGGGCGACGGAACGATCCAGAGACACTGGAATGCTTGGATGGCTG 899
Db 844 CTTTGGACGCTTCGCGGCCGTCGAGAAGATCACTCCACTATTGAGGGATTGGAGATGATC 903
Qy 900 AGGAGCTGTGGAGCTATTGCCGTCACGAGGGTCTGTGTCACCTGCGCACTGCTTT 959
Db 904 CGAGAACTTTGGGCTCTCCTCAACATGTCTCTCTGTTGGCGCCTTACCCCTGAGCCTTAC 963
Qy 960 TAGAGCGGGTACCTCACCGAGGAGCAGCTCCCTTCAATATAAATTTTACCTTGTCT 1019
Db 964 TTGCCCCCTGGCGCCATTCCTATGAGCACTTCCGTCATGCGCAATTACACCTTACAC 1023
Qy 1020 CCTAAAGCCGACTTCCCGAACCAAGATCTATTTCCTGCTTTTGGGAGAGACGACAAA 1079
Db 1024 CATAATGATCCCATACAGAACCGCAAGTACTTACTTGTGTTCGGCATGAATGATG 1083
Qy 1080 ACCATCGGAGGATTTGGCCACTCTTTTGAGCAGAGGTTGGGTCGCTGGCTAAG 1139
Db 1084 GAGGTGACTAATGCACTCAGCATCTTTCATGAGGATGAATGAGGCGATATGGCAATG 1143
Qy 1140 AGGTATCCAGCGGATTTGGCATCTACTATPCCGATGTGGACCTGCGAGCCGCAATCAC 1199
Db 1144 AAATACAAAGCCTGCGCTCAGGGAATCTTCCCGCATCATGATTACGAGCCCTGATTTAT 1203
Qy 1200 CTCGAGCGGTGATCTCTTCTTCAAGGGGAAAACCGTACATGATGTGTACTTC 1259
Db 1204 ATCCACTCGTACATTTCCCTTCTCTCCGAAAGCAAGCCATATTAAAGTGTGTATCTC 1263

Qy 1260 CATACCTTCGAAGC 1273
Db 1264 CACTCCTTGAAC 1277
RESULT 5
ABK15521
ID ABK15521 standard; DNA; 1353 BP.
XX
AC ABK15521;
XX
DT 08-MAY-2002 (first entry)
XX
DNA encoding dimethylallyltryptophan synthase (DmaW) version #2.
DE Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
KW Balansia; Balansiolepis; Echinodopsis; Atkinsonella; Myriogenospora;
KW Neotyphodium; Parepichloe; knockout; antisense technology; gene; ds.
XX Neotyphodium coenophialum.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1353
FT /*tag= a
FT /product= "dmaW"
FT /note= "Dimethylallyltryptophan synthase"
XX
FN US6335188-B1.
XX
XX 01-JAN-2002.
XX 03-MAR-2000; 2000US-00518657.
XX 22-MAR-1999; 99US-0125490P.
XX (KENT) UNIV KENTUCKY RES FOUND.
XX Schardl CL, Wang J;
XX WPI; 2002-163205/21.
P-P5DB; AAU76413.
XX New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW molecule) from fungi that are symbionts of commercially important grasses, useful to engineer ergot alkaloid-deficient symbionts.
XX
XX Claim 1; Col 13-16; 16pp; English.
XX
XX The invention describes an isolated dimethylallyltryptophan synthase (dmaW) nucleic acid (I). (I) is a dmaW sequence from epiphyte fungi Neotyphodium coenophialum symbiont of commercially significant grasses and is useful for expressing dmaW in a cell by recombinant techniques. (I) is also useful for: identifying endophytes e.g. from commercial plants used in forage, pasture, turf, land reclamation and soil conservation that contain or lack a dmaW gene and producing increased amount of ergot alkaloids involving expressing (I) in a host fungal cell so that the copy number of mRNA derived from transcription of the nucleic acid molecule is increased, allowing the host fungal cell to grow under appropriate growth conditions, thus causing increased production of ergot alkaloid. Probes derived from (I) are useful for identifying Neotyphodium or Acremonium or Epichloe that lack dmaW and therefore are unlikely to produce ergot alkaloids. The primers are useful for amplifying segments of dmaW from fungi in family Clavicipitaceae. (I) is also useful for identifying related sequences such as from Balansia, Balansiolepis, Echinodopsis, Atkinsonella, Myriogenospora, Neotyphodium, and Parepichloe, or natural or induced mutants. A knockout construct of (I) or antisense construct is useful for engineering ergot alkaloid-deficient fungal symbionts (endophytes of plants). This sequence encodes a dimethylallyltryptophan synthase (dmaW) described in the method of the invention

XX	Sequence	1353 BP; 373 A; 345 C; 279 G; 356 T; 0 U; 0 Other;	
XX	Query Match	14.7%; Score 205.2; DB 6; Length 1353;	
XX	Best Local Similarity	50.8%; Pred. No. 2.4e-54;	
XX	Matches	647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;	
QY	33	GCAGCACTGCTGCGCAAGCCCTTCTAGCTGCTGAGTCAAGCCCTGACCTCTCGAAC	92
DB	10	GCAAGACATCTCCACAGGAAGTTTATCAACCCCTCAGCGAACAATTTGACCTTGGCAAC	69
QY	93	AAGGACACACAAATAGTGTATAGCACAGCTCCGATGTTTCCACCATGATGGCGGG	152
DB	70	AATGACCAGAGGCTATGCTGGCAGCAGCGCGCAATGTTCCAAAGATACCTCCAACT	129
QY	153	GCCGGCTATGATGTTACGACACAGTACAGTTCTCTGTATCCACCGTGAGGTCAATC	212
DB	130	GCTAACTATAGCATTTATGCTCAGTATCAACATCTGAGCATTTATAAAGCCATATCAT	189
QY	213	CCGCGCTTGGGTCCATACCCAGAAAGGCTCAGCCATGCACTGGAAGAGTCACTCACA	272
DB	190	CTTTTCTTGGTGTCTATCTCAGAGAGTGG---CGAGCGGTGGCTAAGCATTTCTTAG	246
QY	273	CGCTTCGAGCTTCGTTTCGAGCTGAGTTCGAATTAATCTCAAACTACTACTAGTGTGCA	332
DB	247	AGATACGGAACCCCGTTTGAAGCTTAATGCTGCTGACTCCATAGTTCGGTATACA	306
QY	333	TTGAGCGCCCTCGTTCCCTGACGGGAAACGAGGATGATCCATTCAACACCCAGGCAATC	392
DB	307	TACGAGCTTAATACGCGCAACTGSCAGCCATCTGATCCGTTCAACACTTTCGTATC	366
QY	393	AGGCTGTTTCCAGGACCTCAAGGCCATGTTTCCAGGCTTGACCTGGAATGGTTTCAT	452
DB	367	TGGAGGCTCTAAAAAGCTTATAGATTTCCAGCCAGGATAGACCTTCAATGGTTTTC	426
QY	453	CATTTCATAAGCATTTGTCGTTTCGGAGGAGGCTCGGACTCTGCTAGATCGAGT	512
DB	427	TACTTTAAACAGAGCTTACACTTGACGCAACAGTCCAGTACCTGCACTCTCAAAAC	486
QY	513	ATTGAGATCCCGTCTTCAAGACACAGAACAACTGGCAGCCGATCTGGAGCATCTGGC	572
DB	487	TTGGTCAAGGAACAGATCAAACTCAAAACAGCTAGCTTACAGCTTAA---AGTGAAC	543
QY	573	GATATTGCTTGAAGACCTACATCTACCGCGGATCAAGTCGATCGGACCGGACCCCA	632
DB	544	AGTTCGTACTCAAGACCTACATCTACCCGGAATGAAGTCCGTGCAACTGGTAAATCG	603
QY	633	AAAGAGAGACTCATGTTTGAAGCAAT---CAAGGCTGCGGACAAAGTTTGGCAAGTTGCC	689
DB	604	GTCCAGGAGCTTGTTGGTTCGCTCCGTCAGCTAGCCAGCAGATGAAGATATCCGT	663
QY	690	ACTCCACTGCAATCTCGAGGAGTTTAT--AGCTGAGGAGACCCACCTCTCGG--	745
DB	664	CCTGCTTTGAAATGTAGAGACTATGTCAGTCTCGCAATAAAGTCCCTACCAAGGAT	723
QY	746	-----CCACTTCTCTCATGCGATTGGTTCAGCGCTCC	779
DB	724	GACAGTCACATACTCCATTAATCTTCAAGCTTCTCTTGGACCTGTGAGTCTTACC	783
QY	780	GAGTCCGGAATCAAGTCTACTGTATGGAACGCCAGCTCGACCTGGCTTCCATCGAAGT	839
DB	784	AGTCTCGTCAAGATCTACCTCTTGAAGCAAGTGTCTCGTTGCGAGGATGGAAGAT	843
QY	840	ATTGAGCTCTCAACGGGCGAGGACGATCCAGAGACACTGGATGGTCTGGATCGGCTG	899
DB	844	CTTTGAGCGCTTGGCGCGCTGAGAGATAGTCCATTAATGAGGATGAGATGATC	903
QY	900	AGGAGCTGTGAGCATTTGCGCGTCAAGGAGGCTGTGTCCACTGCGCAACTGCTTT	959
DB	904	CGAAGACTTTGGGTCTCTTAACATGTCTCTGGTTTGGCGGCTTACCTGAGCTTAC	963
QY	960	TACGAGCGGCTACCTCAGCGAGGAGCTCCCTTCAATTAATTAATTTACCTTGTCT	1019

DB	964	TTGCCCTCGCGCCATTCCTCAATGAGCAACTTCCGTCCATGGCAATTACACCTTACAC	1023
QY	1020	CCTAAAGCGCACTTCCGGAACCAAGATCTATTTCCCTGCTTTTGGGAGAGCAACAA	1079
DB	1024	CATAATGATCGATACCAAGACCGCAAGTGTACTTTACTGTGTTCCGCAATGATATG	1083
QY	1080	ACCATCGGGAAGGATGGCCACCTTCTTTGAGAGCAGAGGTTGGGTGGCTTGGCTAAG	1139
DB	1084	GAGTGACTAATGACTCAAGAAATTTCTCATGAGGATGAATGAGCGATATGGCAAGT	1143
QY	1140	AGTATCCAGCGGATTTGGCATCTTACTATCCGATGTGACCTGCGAGACCGCAATCAC	1199
DB	1144	AAATACAAAGCTGCTTAGGAAATTTTCCCGCATATAATACGAAGCCCTAAATAT	1203
QY	1200	CTGAGGCGGATCTCTTCTTCAAGGGGNAACCGTACATGAGTGTGTACTTC	1259
DB	1204	ATCCACTGCTACATTTCTTCTCTCTCCGAAATAACAAAGCCATATTTAAGTGTATCTC	1263
QY	1260	CATACCTTCGAAGC 1273	
DB	1264	CACATTTGAAC 1277	

RESULT 6
 ABK15523
 ID ABK15523 standard; DNA; 1598 BP.
 XX
 AC ABK15523;
 DT 08-MAY-2002 (first entry)
 XX
 DE DNA encoding dimethylallyltryptophan synthase (DmaW) version #2.
 XX
 KW Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 KW ergot alkaloid; Neotrophodum; Ascremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansiopepsis; Echinodothis; Atkinsonella; Myriogenospora;
 KW Neotrophodum; Parepichloe; knockout; antisense technology; gene; ds.
 OS Neotrophodum coenophialum.
 PH Key Location/Qualifiers
 FT CDS 25..1511
 FT /*tag= a
 FT /product= "dmaW"
 FT /note= "Dimethylallyltryptophan synthase. This sequence
 contains introns, the positions of which are not detailed
 in the specification"
 XX
 PN US6335188-B1.
 XX
 PD 01-JAN-2002.
 XX
 PF 03-MAR-2000; 2000US-00518657.
 XX
 PR 22-MAR-1999; 99US-0125490P.
 XX
 PA (KENT) UNIV KENTUCKY RES FOUND.
 XX
 FI Schardl CL, Wang J;
 XX WPI; 2002-163205/21.
 DR P-PSDB; AAU76413.
 XX
 PT New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 PT molecule) from fungi that are symbionts of commercially important
 PT grasses, useful to engineer ergot alkaloid-deficient symbionts.
 PS Disclosure; Col 19-20; 16pp; English.
 XX
 CC The invention describes an isolated dimethylallyltryptophan synthase
 CC (dmaW) nucleic acid (1). (1) is a dmaW sequence from epiphyte fungi
 CC Neotrophodum coenophialum symbiont of commercially significant grasses

CC and is useful for expressing DnaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a DnaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack DnaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of DnaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansiolepis, Balansiolepis,
 CC Echinodothia, Atkinsonella, Myriogonospora, Neotyphodium, and
 CC Parepilachne, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This sequence encodes a
 CC dimethylallyltryptophan synthase (dmat) described in the method of the
 CC invention
 XX
 SQ Sequence 1598 BP; 453 A; 392 C; 330 G; 423 T; 0 U; 0 Other;
 Query Match 12.7%; Score 176.6; DB 6; Length 1598;
 Best Local Similarity 50.7%; Pred. No. 3.7e-45;
 Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;
 33 GCAGCAACTGTGTCGCAAGCCCTTCTAGGTGCTGAGTCAGTCAGCCCTGAACTCTCGAAC 92
 34 GCAAGACATCTCCACGAGGAAGTTTATCAACCCCTCAGCGAACAATTTGACTTTGCCAAC 93
 93 AAGGACACACAAATAGTGTATGACACAGCTCCGATGTTGCCACCATGATGGCGGG 152
 94 AATGACAGAGGTATGTTGGCAGACAGCGGGCCAAATGTTCCAAAGATATCCAACT 153
 153 GCGCGTATGATGTTCAAGCAGACAGTAAAGTTCCTCTGATCCACCGTGAAGTATCATC 212
 154 GCTAATATAGCATTTATGTCAGTATCAACATCTGAGCATTTATAAAGCCATATCAT 213
 213 CCGCGGTGGTCCATACCCAGAAAGGTCAGCCCATGACCTGGAAGTATCATCTACA 272
 214 CTTTCTGTTGCTGCTATCTACAGAGTGG---CGAGCGGTGGCTAAGCATTTCTACG 270
 273 CGCTTCGGAGTTCCTTTTCAGAGTTCAGTTCATTAATCTCAAAATCACTACTACGGTTTGA 332
 271 AGATACGGAACCCGGTTGAGCTAAGTCTTAATGCTCTGACTCCATAGTTCGGTATACA 330
 333 TTGAGCCCTCGGTTCCCTGACGGGAAAGAGATGATCCATTAACACCCAGGCAATC 392
 331 TAGGAGCTATTAAACCGCAACTGGCAGCCATCTCGATCCGTTCAACACTTTCGCTATC 390
 393 AGGCTGTTCTCAGGACCTCAAGGCCATGGTTCCAGGGCTTGACCTGGAATGGTTCCGAT 452
 391 TGGAGGCTCTAAAAGCTTATAGATTCCAGCCAGGCAATAGACCTTCAATGGTTTCC 450
 453 CATTTCACTAAAGATTGGTGTTCGGAGGAAGAGGCTCGGACTCTGCTAGATCGAGAT 512
 451 TACTTTAAACAGAGCTTACACTTGACGGAACAGAGTCCACGTCACCTGCACTCTCAAAAC 510
 513 ATTGAGATCCCGTCTTCAGACACAGCAAACTGCGAGCCGATCTGGAGCCATCTGGC 572
 511 TTGCTCAGGAACAGATCAAAATCTAAACAACTAGCTAGCTTAA---AGTGAC 567
 573 GATATTGCTTGAAGACTACATCTACCCCGGATCAAGTCGATCGGACCGGAGCCCA 632
 568 AAGTTGCTACTCAAGACCTACATCTACCCGGAATTTGAAGTCCGTCGCAACTGTAATCG 627
 633 AAGAGAGATCTATGTTTGAAGCAAT---CAAGGCTGCGACAGAAAGTTGCAAGTTGCC 689
 628 GTCCAGGAGCTTGTGTTGGCTCCGTCGCAAGCTAGCGCAGAGATAGATATCCGT 687
 690 ACTCCACTGCAATCTCTCGAGGAGTTTAT---AGCTGAGGAGCAGCCACCTCTCTCGG-- 745
 688 CTTGCTTTGAAATGCTAGAGACTATGTCAGTCTCGCAATTAAGTCCCTACACGGAT 747

QY 746 -----CCACTTCTCTCATGCGATTTGGTCAAGCCGTCC 779
 DB 748 GACAGTCACAATACCTCCATTATCTTACGCGCTTCTCTCTTCCGACCTGGTGAAGTCTTACC 807
 QY 780 GAGTCCCGAATCAAGGTCTACTGTATGGAACGCCAGCTCGACCTGGCTCCATCGAAGGT 839
 DB 808 AAGTCTCGTCAAGATCTACCTCTGGAACGAATGTCTCGTTGCCAGGATGGAAGAT 867
 QY 840 ATTTGGAATCTCTCAACGCGGACGGAACGATCCAGAGACACTGGATGGTCTGGATGCGCTG 899
 DB 868 CTTTGGAGCGCTTGGCGGCCGCTCGAAGAGATCAGTCCACTATTGAGGATTTGGAGATGATC 927
 QY 900 AGGAGCTGTGGAGCTATTGGCCGTCAGGAGGCTGTGTCTCACTGCGCAACTGCTTT 959
 DB 928 CGAGAACTTTGGGCTCTCCCTTAACATGCTCTCTGGTTGGCGGCTTACCTCGACCTTAC 987
 QY 960 TACGAGCGGGTACTCTCACGCGAGGAGCAGCTCCCTTCAATATAAATTTTACCTTCTCT 1019
 DB 988 TTGCCCCCTCGCGCCATTCCCAATGAGAACTTCCGTCCTGCGCAATTACACCTTACAC 1047
 QY 1020 CCTAAAAGCGCACTTCCCGAACCAAGATCTATTTCCTCTGTTTGGGAGAGACGACAAA 1079
 DB 1048 CATAATGATCCGATACCAAGAACCGAAAGTGTACTTTACTGTGTTCCGCATGATGATG 1107
 QY 1080 ACCATCGCGGAAGATTGGCCACCTTCTTTGAGAGCAGAGGTTGGGCTGGCTTGGCTAAG 1139
 DB 1108 GAGTGACTAATGACTCAAGAAATTTCTCATGAGGATGAATGAGGATATGGCAAGT 1167
 QY 1140 AGCTATCCAGC 1150
 DB 1168 AATACAAAGC 1178
 RESULT 7
 AAF11858
 ID AAF11858 standard; cDNA; 676 BP.
 XX
 AC AAF11858;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:4381.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US007781.
 XX
 XX 22-MAR-1999; 99US-00273623.
 XX
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 PI
 XX WPI; 2000-594572/56.
 DR
 XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags.
 XX
 PS Claim 88; Page 1875; 3161pp; English.
 XX

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring the
 CC global expression of genes from FF cells allows the production potential
 CC of the microorganisms to be improved. New genes may be discovered,
 CC possible functions of unknown open reading frames can be identified and
 CC gene copy number variation and stability can be monitored. The expression
 CC of genes can be used to study how FF cells adapt to changes in culture
 CC conditions, environmental stress, spore morphogenesis, recombination,
 CC metabolic or catabolic pathway engineering. Using ESTs provides several
 CC advantages over genomic or random cDNA clones including elimination of
 CC redundancy as one spot on an array equals one gene or open reading frame,
 CC and organization of the microarrays based on function of the gene
 CC products to facilitate analysis of the results. AAF07478 to AAF11247
 CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents
 CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from
 CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from
 CC *Trichoderma reesei*, which are all specifically claimed in the present
 CC invention
 XX
 SQ Sequence 676 BP; 178 A; 156 C; 167 G; 174 T; 0 U; 1 Other;

Query Match 12.6%; Score 175.6; DB 3; Length 676;
 Best Local Similarity 57.8%; Pred. No. 4.7e-45;
 Matches 351; Conservative 0; Mismatches 250; Indels 6; Gaps 2;

QY 52 AGCCCTCTACGTGCTGAGTCACAGCCCTGACCTCTCGAACAGGACACACAAATGCT 111
 DB 70 AACATTCAGGTTCTCAGCAAGATCTTAGATTTTCGCAACGAGGACGAGACTGCT 129
 QY 112 GGTATAGCAGCTCCGANTTTGCGACCATGATGCGGGGCGGCTGATGTTTCAG 171
 DB 130 GGCACAGTACCGGCCAATGTATGCCAAATCTTAAAGGATGCGGCTACGCAATCATG 189
 QY 172 CACAGTACAAAGTTCCTCTGATCCACCGTCAGGTGATCATCCGGGCTGGTCCATACC 231
 DB 190 CCATATCTATATCTCTGCTCCACACAAATGTGTGTTCCGATCTGGGCCCTATC 249
 QY 232 CAGAAAGGCTCAGCCCATCTCACTGGAAGATCATCTCAGCGTTTCGGACTTCCTTTG 291
 DB 250 CGGGAACGCGCAGAGAC---CGATGATGAGCATCTCAGCGGATTCGGTCTACCTTATG 306
 QY 292 AGCTGAGCTCAATTAATCTCAATCACTACTAGGTTTCGATCGAGCCCTCGGTCC 351
 DB 307 AATTGAGCTGAATGCTCTAATTCGGTCTGGGTTTCGATTCGAGCCCATTCGACCTT 366
 QY 352 TGACGGGACCAAGAGATGATCCATTCAACACCGAGCAATCAGGCTGTCTCCAGGACC 411
 DB 367 TATCAGGACTGAGCAGACCCCTTCATGACATGTCATCTGGGATCTCTTGGGAAC 426
 QY 412 TCAAGCCCATGTTTCAGGGCTGACCTGGAATGTTTGCATCATTTCACTAAGCAATGG 471
 DB 427 TCCGGAAGCTGGGTTCTGATTTTGATCTCCAGTGGTTTCCCAATTCAGAGGATCTGG 486
 QY 472 TCGTTTCGGAGGAAGAGGCTCGGACTCTGCTAGATCGAGATATGAGATCCCGCTTCA 531
 DB 487 TTTTAGATCGAGGAAACGAAGTTGTTTAGAGACACACGGTCTGACAGGACAGGTTA 546
 QY 532 AGACAGACAAACTGGCAGCGGATCTCGAGCCATCTCGGATATGTTCTTGAAGACT 591
 DB 547 AGACTCAGAACAACTANGCGTGGACTTCAA---AGGTGGAATAATTCGAGGTCAAGATGT 603
 QY 592 ACATCTACCGGGATCAAGTGCATCGGACCGGACCCCAAGAGAGACTCATGTTTG 651
 DB 604 ACATGATCTTATCTGAAATCGTTGCTACCGGATCCCAATTGAGGACTCATGTTTG 663
 QY 652 ACGCAAT 658

DB 664 ATTCGAT 670

RESULT 8

ABK15522

ID ABK15522 standard; DNA; 1908 BP.

XX ABK15522;

AC ABK15522;

DT 08-MAY-2002 (first entry)

XX DNA encoding dimethylallyltryptophan synthase (DmaW) version #1.

DE Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 XX endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansioopsis; Echinodothis; Atkinsonella; Myriogenospora;
 KW Neotyphodium; Parepichloe; Knockout; antisense technology; gene; ds.
 XX Neotyphodium coenophialum.

OS

FF Key Location/Qualifiers

FT CDS 277..1756

FT /*tag= a

FT /product= "dmaW"

FT /note= "Dimethylallyltryptophan synthase. This sequence
 contains introns, the position of which are not detailed
 in the specification"

XX

PN US6335188-B1.

XX 01-JAN-2002.

PD

XX 03-MAR-2000; 2000US-00518657.

PF

XX 22-MAR-1999; 99US-0125490P.

PR

XX (KENT) UNIV KENTUCKY RES FOUND.

PA

XX Schardl CL, Wang J;

PI

XX WPI; 2002-163205/21.

DR

XX P-ESDB; AAU76412.

DR

XX New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 molecule) from fungi that are symbionts of commercially important
 grasses, useful to engineer ergot alkaloid-deficient symbionts.

PT

XX Disclosure; Col 17-20; 16pp; English.

PS

XX The invention describes an isolated dimethylallyltryptophan synthase
 (dmaW) nucleic acid (I). (I) is a DmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses
 CC and is useful for expressing DmaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a DmaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack DmaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of DmaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansioopsis,
 CC Echinodothis, Atkinsonella, Myriogenospora, Neotyphodium, and
 CC Parepichloe, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This sequence encodes a
 CC dimethylallyltryptophan synthase (dmaW) described in the method of the
 CC invention

SQ	Sequence	1908 BP; 544 A; 449 C; 395 G; 520 T; 0 U; 0 Other;	
	Query Match	12.4%; Score 173.4; DB 6; Length 1908;	
	Best Local Similarity	50.5%; Pred. No. 4.3e-44;	
	Matches	581; Conservative 0; Mismatches 531; Indels 39; Gaps 5;	
QY	33	GCAGCAACACTGCTGCGCAAGCCCTTCTAGCTGCTGAGTCAAGCCCTGAACTCTCGAAC	92
DB	286	GCAAGACACTCCACAGAGATTATCATACCTTAGCGAAACATTTCACCTTGGCAAC	345
QY	93	AAGGACACACAAAATGGTGGTATAGACACAGCTCGATGTTTGGCCACCATGATGGGGG	152
DB	346	AATGACAGAGGCTATGGTGGCAGACAGCGGGCCCAATGTCGAAAAGATGCTCCAACT	405
QY	153	GCGGGTATGATGTTCCACGACAGTACAAGTTCCTCTGTATCCACCGTGAGTCAATC	212
DB	406	GCTAACTATAGCAATGATGCTCAGTATCGACATCTCGGCATTTTAAAGGCCATGTCAIT	465
QY	213	CCGGCGTTGGTCCATACCCAGAAAAGGTCAGCCCATGCACCTCGGAAGATCATCTCACA	272
DB	466	CCTTTCTTGGTCTATCTCTACAGAAGTGG--CGAGCGGTGGCTAAGCATCTTACG	522
QY	273	CGCTTGGACTTCCTTTGAGCTGAGCTTCAATTAATCTCAATCACTACTAGCTTGA	332
DB	523	AGATACGGAAACCCGTTTGAAGTAACTTAAATGTCTCGGACTCGGTATTCGGTATACA	582
QY	333	TTGAGCCCTCGGTTCCCTGACGGGAAAGAGATGATCCATTCACACCCAGGCAATC	392
DB	583	TAGAGCCTATTAAACGCGCACTGCGAGTCTATCTCGATCCGTTTAACTTTGCGTATC	642
QY	393	AGGCGCTGTTCTCAGGACCTCAAGGCCATGTTTCCAGGGCTTGAACCTGGAATGTTGAT	452
DB	643	TGGAGGCGCCCTGAAAAGACATATTGAGTCCAGCCAGGCATAGACCTTGAATGGTTTCT	702
QY	453	CATTTCACATAACGATTGTCGTTTCGGAGGAGAGGCTCGGACTCTGCTAGATCGAGAT	512
DB	703	TACTTTAAACAGAGCTTACACTTTAGCGCAACAGATCCAGTACCTGCACTCGCAAC	762
QY	513	ATTGAGATCCCCGTTCTTCAAGACACAGAACTCGCAGCCGATCTGGAGCCATCTGCG	572
DB	763	TTGTTTAAAGAAACAGATCAAACTCAAACTCAAACTCGTTTGGACCTTAA--AGTGAC	819
QY	573	GATATGCTTTGAAGACCTACATCTACCCGCGATCAAGTCAATCGGACCGGACCCCA	632
DB	820	AACTTCGTACTGAGACCTACATCTACCCGCAATGAAGTCCGTCGCAACTGGTAAATCG	879
QY	633	AAAGAGAGACTCATGTTTGAAGCAAT--CAAGGCTGCGCAAGATTGGCAAGTTGCG	689
DB	880	GTCCAGAGCTCGTGTGCTCGTCCGCAAGCTAGCGCAGAGCAAGATATCCGT	939
QY	690	ACTCCACTGGCAATCTCGAGGATTTAT--AGCTGAGGAGCAGCCACCCCTCTCGG--	745
DB	940	CCTGCTTTGAAATGCTAGAAGACTATGTCCAGTCTCGCAATAAATCTCTACACCGAT	999
QY	746	-----CCACTTTCTCTCATGCAATTTGGTCAAGCCGCTG	779
DB	1000	GACAGTCAAACTACTCTATTATCTTACGCTTCTCTTGGCACTGATAGTCTCTACC	1059
QY	780	GAGTCCCGAATCAAGGTCTACTGATGGAACGCCAGCTGACCTGCGCTCCATCGAAGT	839
DB	1060	AAGTCTCGTCAAGATCTACTCTCGGAACGAATGCTCTCGTTGCGCAGCATGGAAGAT	1119
QY	840	ATTGAGTCTCAACGCGCAAGCAATCCAGACACTGATGCTGATGCTGGTGGCTG	899
DB	1120	CTTTGGACGCTTGGCGGCGCTGAGAGATCAGTCCACTATTGAGGATTTGGAGATC	1179
QY	900	AGGAGAGCTGGGAGCTATTGCGCGTCAGCGAGGGTCTGTGTCCACTGCGCAACTGCTTT	959
DB	1180	CGAGACTTTGGGGTCTCTCAACATGCTCTCTGTTTGGCGCTTACCTGAGCCTTAC	1239
QY	960	TACAGCGCGGTACTCAGCGAGGAGCAGCTCCCTCTGATTAATAATTTACCTTGCT	1019
DB	1240	TTGCCCTCGGCGCAATTCCTCAATGAGCAACTTCCGTCCTGCAATGACCTTTACAC	1299

QY	1020	CCTAAAGCGCACTTCCGACACACAGATCTATTTCCTCTGCTTTTGGGACAGCAAA	1079
DB	1300	CATAATGATCCATACAGAACCGCAAGTGTACTTTTACTGTGTCGCGCATGATGATG	1359
QY	1080	ACCATCGCGGAAGATTGGCCACCTTCTTTAGAGCAGAGGTTGGGTGGCTTGGCTAAG	1139
DB	1360	GAGGTGACTAATGCACTCAGACATTTCTCATGAGGCATGAATGCGCATATGCGCAAGT	1419
QY	1140	AGCTATCCAGC	1150
DB	1420	AAATACAAAGC	1430
RESULT 9			
AAF12651			
ID	AAF12651	standard; cDNA; 657 BP.	
XX	AC	AAF12651;	
XX	DT	13-MAR-2001 (first entry)	
XX	DE	Aspergillus oryzae EST SEQ ID NO:5174.	
XX	DE	Multiple gene expression; filamentous fungal cell; EST;	
KW	KW	expressed sequence tag; Fusarium venenatum; Aspergillus niger;	
KW	KW	Aspergillus oryzae; Trichoderma reesei; identification; recombination;	
KW	KW	culture condition; environmental stress; spore morphogenesis;	
KW	KW	metabolic pathway engineering; catabolic pathway engineering; ss.	
XX	OS	Aspergillus oryzae.	
XX	XX	WO200056762-A2.	
XX	XX	28-SEP-2000.	
XX	XX	22-MAR-2000; 2000WO-US007781.	
XX	XX	22-MAR-1999; 99US-00273623.	
PA	PA	(NOVO) NOVO NORDISK BIOTECH INC.	
PA	PA	(NOVO) NOVO NORDISK AS.	
PI	PI	Berka RM, Rey WM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;	
XX	XX	WPI; 2000-594572/56.	
PT	PT	Monitoring differential expression of genes in filamentous fungal cells	
PT	PT	uses fluorescence-labeled nucleic acids isolated from the cells and a	
PT	PT	substrate of expressed sequence tags.	
XX	XX	Claim 88; Page 2165; 3161pp; English.	
XX	XX	The present invention describes a method for monitoring differential	
CC	CC	expression of genes in a first filamentous fungal (FF) cell relative to	
CC	CC	expression of the same genes in one or more second filamentous fungal	
CC	CC	cells. The method uses fluorescence-labeled nucleic acids isolated from	
CC	CC	the FF cells and a substrate of expressed sequence tags (EST). The ESTs	
CC	CC	are used in the methods for monitoring differential expression of genes	
CC	CC	in a first filamentous fungal (FF) cell relative to expression of the	
CC	CC	same genes in one or more second filamentous fungal cells. Monitoring the	
CC	CC	global expression of genes from FF cells allows the production potential	
CC	CC	of the microorganisms to be improved. New genes may be discovered.	
CC	CC	possible functions of unknown open reading frames can be identified and	
CC	CC	gene copy number variation and stability can be monitored. The expression	
CC	CC	of genes can be used to study how FF cells adapt to changes in culture	
CC	CC	conditions, environmental stress, spore morphogenesis, recombination,	
CC	CC	metabolic or catabolic pathway engineering. Using ESTs provides several	
CC	CC	advantages over genomic or random cDNA clones including elimination of	
CC	CC	redundancy as one spot on an array equals one gene or open reading frame,	
CC	CC	and organization of the microarrays based on function of the gene	
CC	CC	products to facilitate analysis of the results. AAF07478 to AAF11247	
CC	CC	represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents	

XX Novel Cynomolgus Monkey Calcium-Sensing Receptor polypeptides are useful
PT for treating infections, pain, cancer, diabetes, obesity, asthma,
PT schizophrenia, hypertension, urinary retention, Parkinson's disease and
PT stroke.
XX
XX Claim 1; Page 27-28; 38pp; English.
XX
CC The sequence represents the coding sequence of Cynomolgus monkey calcium
CC -sensing receptor. The calcium-sensing receptor nucleic acid and
CC polypeptide are useful for treating bacterial, fungal, protozoan and
CC viral infections, particularly infections caused by HIV-1 or HIV-2, pain,
CC cancer, diabetes, obesity, anorexia, bulimia, Parkinson's disease, acute
CC heart failure, hypotension, hypertension, urinary retention,
CC osteoporosis, angina pectoris, myocardial infarction, stroke, ulcers,
CC asthma, allergies, benign prostatic hypertrophy, migraine, vomiting,
CC psychotic and neurological disorders including anxiety, schizophrenia,
CC manic depression, depression, delirium, dementia and severe mental
CC retardation, and dyskinesias, such as Huntington's disease or Gilles de
CC la Tourette's syndrome. The polypeptide is useful for structure-based
CC design of agonist, antagonist or inhibitor of the calcium-sensing
CC receptor. The DNA sequence is useful for chromosome localisation studies.
CC The polypeptide and nucleic acid are also useful as vaccines. (Updated on
CC 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 3237 BP; 746 A; 923 C; 824 G; 744 T; 0 U; 0 Other;

Query Match 2.5%; Score 34.8; DB 4; Length 3237;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 114; Conservative 0; Mismatches 112; Indels 2; Gaps 1;
773 GCGCTCGAGTCCGGAATCAAGGTCTACTGTATGGAAGCGCAGCTCGACCTGCCCTCCAT 832
Db 3009 GCITTTCTGGGCTCCAGGAGTCTCGTGTGTAGATTCCTGTGGGCAATGCGTCTT 2950
833 CGAAGGTATTGGACTCTCAACGGGGGACGGAACGATCCAGAGAC--ACTGGATGCTCTG 890
Db 2949 CTGAGGCTCATCGAAGCTCAGTGAGAGGTGACCGTGCACCTGCCGAAGATGACCTTCTG 2890
891 GATGCGCTGAGGAGCTGTGCGAGCTATTGCCCGTCAACGAGGCTGTGTCCACATGCCG 950
Db 2889 CTTCGATCTGGCTGCTGCTGCGATTTCTGCTGTGGAGGTGAGGGCTGCTGCTG 2830
951 AACTGCTTTTACGAGCCGGTACTCTACCGCAGGAGCAGCTCCCTTC 998
Db 2829 CTGCTCTTGTGGTGTAGGGCCAGCGCTCTGCTGCTTCTGCTCTC 2782

RESULT 15
ACA42624
ID ACA42624 standard; DNA; 4254 BP.

XX ACA42624;
XX
XX 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #24281.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Pseudomonas aeruginosa.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.

06-VAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.
P-PSDB; ABU38754.

New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 30494; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway;
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Sequence 4254 BP; 837 A; 1404 C; 1389 G; 624 T; 0 U; 0 Other;

Query Match 2.5%; Score 34.8; DB 7; Length 4254;

Best Local Similarity 50.0%; Pred. No. 15;

Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

508 GAGATATTGAGATCCCGCTCTCAAGCACAGACAACTGGCAGCCGATCTGGAGCCAT 567

1373 GCGCGAGTGGCATCAGCTCGTGCAGAGCCAGACAGCCCAACATCGCTCGGCGCTCA 1432

568 CTGGCGATATTGCTTGTGAAGACCTACATCTACCCCGGATCAAGTCGATCGCAGCGGA 627

1433 ATGGCGTGACTGCTGTGCAGAGCCAGACGGTGCCAAATATCGTTTCGGCGCAAGTGCA 1492

628 CCCCCAAAGAGAGACTCATGTTTTCGCAATCAAGGCTGCCGACAAAGTTTGCA 681

1493 TCACCGTTGTGCAAGCCAGAAATGCGCAAAATATCGTTTCAGGCGCGAGTGCA 1546

Search completed: July 28, 2004, 22:37:15

Job time : 627 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 22:17:55 ; Search time 121 Seconds
(without alignments)
6388.815 Million cell updates/sec

Title:
Perfect score: 1393
Sequence: 1 ccgaagtgtgagcaatggag.....cgaaaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A-COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B-COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A-COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B-COMB.seq:*
5: /cgn2_6/prodata/2/ina/5A-COMB.seq:*
6: /cgn2_6/prodata/2/ina/5B-COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match %	Length	ID	Description
1	1393	100.0	1393	4	US-09-472-364-1 Sequence 1, Appli
2	205.2	14.7	1347	4	US-09-518-657-1 Sequence 1, Appli
3	205.2	14.7	1353	4	US-09-518-657-3 Sequence 3, Appli
4	176.6	12.7	1598	4	US-09-518-657-6 Sequence 6, Appli
5	173.4	12.4	1908	4	US-09-518-657-5 Sequence 5, Appli
6	37.8	2.7	2830	2	US-09-010-928B-1 Sequence 1, Appli
7	35.2	2.5	339	4	US-09-821-976-8976 Sequence 8976, Ap
8	34.8	2.5	606	4	US-09-252-991A-10059 Sequence 10059, A
9	34.8	2.5	4248	4	US-09-252-991A-9867 Sequence 9867, Ap
10	33.6	2.4	1560	1	US-07-669-171-1 Sequence 1, Appli
11	32.8	2.4	9098	4	US-09-358-082A-28 Sequence 28, Appli
12	32.8	2.4	4403765	3	US-09-103-840A-2 Sequence 2, Appli
13	32.8	2.4	4411529	3	US-09-103-840A-1 Sequence 1, Appli
14	32.6	2.3	2416	1	US-08-592-126-60 Sequence 60, Appli
15	32.6	2.3	2416	4	US-09-168-595-60 Sequence 60, Appli
16	32.2	2.3	12848	4	US-09-453-702B-252 Sequence 252, App
17	32.2	2.3	26173	4	US-09-453-702B-69 Sequence 69, Appli
18	32.2	2.3	38155	4	US-09-453-702B-79 Sequence 79, Appli
19	32.2	2.3	48908	4	US-09-453-702B-137 Sequence 137, App
20	32	2.3	240	1	US-08-209-747-11 Sequence 11, Appli
21	32	2.3	240	1	US-08-458-298-11 Sequence 11, Appli
22	32	2.3	308	4	US-09-621-976-80 Sequence 29, Appli
23	32	2.3	337	4	US-08-868-452-29 Sequence 1, Appli
24	32	2.3	339	1	US-08-486-057B-1 Sequence 29, Appli
25	32	2.3	339	1	US-08-470-837-29 Sequence 1, Appli
26	32	2.3	339	2	US-08-789-588-1 Sequence 1, Appli
27	32	2.3	339	3	US-09-123-233-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-472-364-1
; Sequence 1, Application US/09472364
; Patent No. 6383781
; GENERAL INFORMATION:
; APPLICANT: Christensen, Bjorn Eggert
; APPLICANT: Mollgaard, Henrik
; APPLICANT: Kaasgaard, Svend
; APPLICANT: Lehbeck, Jan
; TITLE OF INVENTION: Methods for producing polypeptides in
; TITLE OF INVENTION: aspergillus mutant cells
; FILE REFERENCE: 4483.200-US
; CURRENT APPLICATION NUMBER: US/09/472.364
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/139,593
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/117,396
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: PA 1999 00745
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: PA 1998 01726
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer 5956
; NAME/KEY: CDS
; LOCATION: (15)...(1328)
US-09-472-364-1

Query Match 100.0%; Score 1393; DB 4; Length 1393;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAGCACTGTCGCAAGGCCCTTCT 60
Db 1 CGGAAGCTGAGCAATGGAGATCTCCAGAAAGCAGCAGCACTGTCGCAAGGCCCTTCT 60

QY 61 AGCTCTGAGTCAAGCCCTGAACTCTCGAAACAGGACCACACAAATGGTGTATAGCA 120
Db 61 AGCTCTGAGTCAAGCCCTGAACTCTCGAAACAGGACCACACAAATGGTGTATAGCA 120

QY 121 CAGCTCCGATGTTTGCCACCATGATGGCGGGGGCGGCGGTATGATGTTTCCGACAGTACA 180
Db 121 CAGCTCCGATGTTTGCCACCATGATGGCGGGGGCGGCGGTATGATGTTTCCGACAGTACA 180

QY 181 AGTTCTCTGATCCACCGTGAAGTCAATCCCGCGGTGGGTGCATACCCAGAAAGG 240
 Db 181 AGTTCTCTGATCCACCGTGAAGTCAATCCCGCGGTGGGTGCATACCCAGAAAGG 240
 QY 241 GTACGCCCATCTCACTGAAGAGTCACTCACAGCTTCGGACTTCCTTTTCGAGCTGAGCT 300
 Db 241 GTACGCCCATCTCACTGAAGAGTCACTCACAGCTTCGGACTTCCTTTTCGAGCTGAGCT 300
 QY 301 TCAATTACTCCAAATCACTACTACGGTTTCATTCGAGCCCTCGGTTCCCTGAAGGAA 360
 Db 301 TCAATTACTCCAAATCACTACTACGGTTTCATTCGAGCCCTCGGTTCCCTGAAGGAA 360
 QY 361 CGRAGGATGATCCATTCACACCCAGGCAATCAGGCTGTTCCAGGACCTCAAGCCA 420
 Db 361 CGRAGGATGATCCATTCACACCCAGGCAATCAGGCTGTTCCAGGACCTCAAGCCA 420
 QY 421 TGGTTCCAGGCTTGACCTGGAATGTTTCGATCATTTTCACTAAGCAATGTTGTTTCGG 480
 Db 421 TGGTTCCAGGCTTGACCTGGAATGTTTCGATCATTTTCACTAAGCAATGTTGTTTCGG 480
 QY 481 AGGAGAGGCTCGACTCTCTAGATCGAGATATTGAGATCCCGCTTCCTCAAGACAGAA 540
 Db 481 AGGAGAGGCTCGACTCTCTAGATCGAGATATTGAGATCCCGCTTCCTCAAGACAGAA 540
 QY 541 ACAAACTGGCAGCGGATCTGGAGCCATCTGGCGATATTGTTTGAAGACCTACATCTACC 600
 Db 541 ACAAACTGGCAGCGGATCTGGAGCCATCTGGCGATATTGTTTGAAGACCTACATCTACC 600
 QY 601 CGCGGATCAAGTGCATCGCAGCGGACCCCAAGAGAGACTCATGTTTGAAGCAATCA 660
 Db 601 CGCGGATCAAGTGCATCGCAGCGGACCCCAAGAGAGACTCATGTTTGAAGCAATCA 660
 QY 661 AGGTCCGCAAGTGTGGCAAGTTGCCACTCCACTCCATCGCAATCTCGAGGAGTTATAG 720
 Db 661 AGGTCCGCAAGTGTGGCAAGTTGCCACTCCACTCCATCGCAATCTCGAGGAGTTATAG 720
 QY 721 CTGAGGAGACCCACCTCTCGGCCATTTCTCATGCGATTTTGGTCAAGCGTCCG 780
 Db 721 CTGAGGAGACCCACCTCTCGGCCATTTCTCATGCGATTTTGGTCAAGCGTCCG 780
 QY 781 AGTCCCGAATCAAGTCTATGTAAGCGGAGCTCGACTGCGCTCCATCGAAGTGA 840
 Db 781 AGTCCCGAATCAAGTCTATGTAAGCGGAGCTCGACTGCGCTCCATCGAAGTGA 840
 QY 841 TTTGGACTCTCAACGGGCGAGCAAGATCCAGAGACACTGGATGTTGGGATGGCTGA 900
 Db 841 TTTGGACTCTCAACGGGCGAGCAAGATCCAGAGACACTGGATGTTGGGATGGCTGA 900
 QY 901 GGGAGCTGGGAGCTATTGCCCGTCAAGGAGTCTGTGCACTGCCGAACTGCTTTT 960
 Db 901 GGGAGCTGGGAGCTATTGCCCGTCAAGGAGTCTGTGCACTGCCGAACTGCTTTT 960
 QY 961 ACCAGCGGGTACTCTACCGCAGGAGCAGTCCCTCTTCAATTAATAATTTTACCTTGTCTC 1020
 Db 961 ACCAGCGGGTACTCTACCGCAGGAGCAGTCCCTCTTCAATTAATAATTTTACCTTGTCTC 1020
 QY 1021 CTAAGAAGCGCACTTCCCGAACCACAGATCTATTTCCTGTTTGGGAGAGACGCAAAA 1080
 Db 1021 CTAAGAAGCGCACTTCCCGAACCACAGATCTATTTCCTGTTTGGGAGAGACGCAAAA 1080
 QY 1081 CCATCCGGAAGGATTTGGCACTCTTTTGAAGACAGAGTTGGGTTGGCTTGAAGA 1140
 Db 1081 CCATCCGGAAGGATTTGGCACTCTTTTGAAGACAGAGTTGGGTTGGCTTGAAGA 1140
 QY 1141 GCTATCCAGCGGATTTGGCATCTACTATCCGATGTGGACCTGCAGACCGCAATCACC 1200
 Db 1141 GCTATCCAGCGGATTTGGCATCTACTATCCGATGTGGACCTGCAGACCGCAATCACC 1200
 QY 1201 TGCAGGCTGGATCTCTTCTCTTCAAGGGGAAAAACCGTACATGATGTGTACCTCC 1260
 Db 1201 TGCAGGCTGGATCTCTTCTCTTCAAGGGGAAAAACCGTACATGATGTGTACCTCC 1260
 QY 1261 ATACCTTCGAAGCGTTTCAGTGTCTGTCGCCAGGAGTGGCTATGTGTCAAGTGGCCACA 1320

Db 1261 ATACCTTCGAAGCGTTTCAGTGTCTGTCGCCAGGAGTGGCTATGTGTCAAGTGGCCACA 1320
 QY 1321 ATCCCTTGGAGCTAGTTTATCCCTTCATCTATGATCGTTCGTTGAATGTGTGGTGAAGAA 1380
 Db 1321 ATCCCTTGGAGCTAGTTTATCCCTTCATCTATGATCGTTCGTTGAATGTGTGGTGAAGAA 1380
 QY 1381 AAAAAAAGGAA 1393
 Db 1381 AAAAAAAGGAA 1393
 RESULT 2
 US-09-518-657-1
 ; Sequence 1, Application US/09518657
 ; Patent No. 6335188
 ; GENERAL INFORMATION:
 ; APPLICANT: Schardl, Christopher L.
 ; APPLICANT: Wang, Jinghong
 ; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
 ; TITLE OF INVENTION: Which Encode Therefor and Related Methods
 ; FILE REFERENCE: P-1060
 ; CURRENT APPLICATION NUMBER: US/09/518,657
 ; CURRENT FILING DATE: 2000-03-03
 ; EARLIER APPLICATION NUMBER: 60/125,490
 ; EARLIER FILING DATE: 1999-03-22
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 1
 ; TYPE: DNA
 ; ORGANISM: Neotyphodium coenophialum
 US-09-518-657-1
 Query Match 14.7%; Score 205.2; DB 4; Length 1347;
 Best Local Similarity 50.8%; Pred. No. 6.5e-57;
 Matches 647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;
 QY 33 GCAGCAACACTGCTGCCAAGCCCTTCTACGTGCTGAGTCAAGCCCTGAACTCTCGAAC 92
 Db 10 GCAAGACACTCCACAGGAGTATTATCATACCTTAGCGAACAATTGACTTTGCCAAC 69
 QY 93 AAGGACCAACAATAATGGTGTATAGCACAGCTCCGATGTTTCCACCATGATGGCGGG 152
 Db 70 AATGACAGAGGCTATGGTGGCACAGCAGCGGCCAATGTTGCAAAAGATGCTCCAACT 129
 QY 153 GCGGCTATGATGTTACGCACAGTACAGTTCCTCTGTATCCACCGTGGGTGATCATC 212
 Db 130 GCTAACTATAGCATTTGATGCTCAGTATCGACATCTGGGCATTTATAAGGCCATGTCATT 189
 QY 213 CCGGCTTGGGTCCATACCCAGAAAGGGTCAAGCCATGACCTGGAAGAGTCACTCATCTCACA 272
 Db 190 CTTTTCTTGGTGTCTATCTCTACAGAGTGG---CGAGCGGTGGCTAAGCATTTCTTACG 246
 QY 273 CGCTTCGAGCTTCCTTTTCGAGCTGAGCTTCAATTAATCTCCAAATCACTACTACGGTTTGA 332
 Db 247 AGATACGGAACCCCGTTTGGCTTAAGTCTTAATTTGCTCGGACTCCGTTAGTTCGGTATACA 306
 QY 333 TTCAGGCCCTCGGTTCCCTGACGGGAAAGAGGATGATCCATTCAACACCCAGGCAATC 392
 Db 307 TAGAGGCTATTAAACCGCGCACTGGCAGTCTCTGATCCGTTTAAACACTTTTCGCTATC 366
 QY 393 AGGCTGTTTCTCAGGACCTCAAGGCCATGGTTCCAGGGCTTGAACCTGGAATGGTTCGAT 452
 Db 367 TGGAGGCCCTGAAAAAGCATATTGAGTCCAGCCAGGCATAGACCTTGAATGGTTTCT 426
 QY 453 CATTTCACTAAAGCATTTGGTCTGTTTCGGAGGAGAGGCTCGGACTCTGCTAGATCGAGAT 512
 Db 427 TACTTTAAACAGAGCTTACACTTTGACGMAACGAGTCCACGTACCTGACTCGCAAAAC 486
 QY 513 ATTGAGATCCCCGTCTTCAAGACACAGAAACAACTGGCAGCGGATCTGGAGCCATCTGCG 572
 Db 487 TTGGTTAAGGAAACAGATCAAAACTCAAAACAGCTCGCTTTTGGACCTTAA---AGGTGAC 543

QY 573 GATATTGCTTGAAGACCTTACATCTACCCCGCGGATCAAGTCGATCCGAGCCCGGACCCCA 632
Db |||||
QY 544 AAGTTGCTACTGAAGACCTACATCTACCCCGAATTGAAGTCCGTCGCAATCTGTAATCG 603
Db |||||
QY 633 AAAGAGAGACTCATGTTTGAACGAAT--CAAGGCTGCCGACAAAGTTTGGCAAAAGTTGCC 689
Db |||||
QY 604 GTCCAGAGAGCTGTTTGGCTCCGTCGCGCAAGTCTAGCGAGAGCAAGAGATACCGT 663
QY 690 ACTCCACTGCAATCTCGAGGAGTTAT--AGCTGAGCGAGACACCCCTCTCGG-- 745
Db |||||
QY 664 CTTGCGCTTTGAAATGCTAGAAGACTATGTCAGTCTCGCAATAAATCTCTACCCAGGAT 723
QY 746 -----CCACTTCTCTCATGCGATTTGCTCAAGCGCTCC 779
Db |||||
QY 724 GACAGTCAAACTCTATTAATCTTCAAGCCCTTCTCTTTCGACCTGATAAGTCTCTACC 783
QY 780 GAGTCCCGAATCAAGGCTTACTGTATGGAACGCCAGCTCGACCTGCGCTCCATCGAAGT 839
Db |||||
QY 784 AAGTCTCGTCAAGATCTACCTCTGGAAGCAAGTGTCTCGTTGCCAGCGATGGAAGAT 843
QY 840 ATTTGAGACTCTAAACCGGCGACGGAAGATCCAGAGACACTGGATGCTGAGTGGCTG 899
Db |||||
QY 844 CTTTGGACGCTTGGCGCGCTCGAGAAAGATCACTGTCACATTTGAGGGATTGGAGATGATC 903
QY 900 AGGAGAGCTGTGGCAGCTATTGCGCGCTCAAGAGGCTGTGTCTCACTGCGCAACTGCTTT 959
Db |||||
QY 904 CGAGAACTTTGGGCTCTCTCAACATGTCTCTGTTTGGCGCTTACCTGAGCCTTAC 963
QY 960 TAGAGCGGCTTACTCAACGAGAGAGCTCCCTTCAATTAATAATTTTACCTTGTCT 1019
Db |||||
QY 964 TTGCGCTCGCGCCATTTCCCAATGAGCAACTTCCGTCATGCGCAATTTACACCTTTACAC 1023
QY 1020 CTTAAAGCGCACTTCCGGAACACAGATCTATTTCCCTGCTTTTGGCGCAGACGACAAA 1079
Db |||||
QY 1024 CATATGATCCATACGAGAACCGAAGTACTTACTGTGTTCGCGATGATGATG 1083
QY 1080 ACCATCGCGAAGATTTGCCACTTCTTTTGAGCAGAGGTTGGGCTGGCTTGGCTAAG 1139
Db |||||
QY 1084 GAGTGACTAATGCACTACGACATTTCTTATGAGCATGAATGGAGCATATGGCAAGT 1143
QY 1140 AGTATCCAGCGATTTGGCATCTTACTATCCCGATGTGAGCTGAGACCGCAATCAC 1199
Db |||||
QY 1144 AATACAAAGCGCTGCTCAGGGAATCTTTCCCGCATCATGATPACAAAGCCCTGAAATAT 1203
QY 1200 CTGAGCGGTGGATCTCTTCTTACAAAGGGGAAAAACCGTACATGATGTGTACCTTC 1259
Db |||||
QY 1204 ATCCACTCGTACATTTCTTCTCTTACCGGGAAGAAACGCAATATTTAAGTGTATCTC 1263
QY 1260 CATACCTTCGAAGC 1273
Db |||||
QY 1264 CACTCCTTTGAAC 1277

RESULT 3

US-09-518-657-3
; Sequence 3, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; APPLICANT: Wang, Jinghong
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; TITLE OF INVENTION: Which Encode Therefor and Related Methods
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; CURRENT FILING DATE: 2000-03-03
; EARLIER APPLICATION NUMBER: 60/125,490
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1353
; TYPE: DNA

; ORGANISM: Neotyphodium coenophialum

US-09-518-657-3

Query Match 14.7%; Score 205.2; DB 4; Length 1353;
Best Local Similarity 50.8%; Pred. No. 6.5e-57;
Matches 647; Conservative 0; Mismatches 588; Indels 39; Gaps 5;

QY 33 GCAGCAACACCTGCTGCCAAAGCCCTTCTACGTGTGAGTCAAGCCTGAACTCTCGAAC 92
Db |||||
QY 10 GCAAGACACTCCACCGAGGAAGTTTATCAAAACCTCCAGCGAAACATTTGACTTTGCCAAC 69
QY 93 AAGGACCACACAAATGGTGTATAGCACAGCTCCGATGTTTGGCACCATGATGGCGGG 152
Db |||||
QY 70 AATGACAGAGGCTATGTTGCGACAGCACGCGCGCAATGTTCCAAAGATACTCCAACT 129
QY 153 GCCGCTATGATGTTACGCAAGTACAAAGTTCCTGTATFCCACCGTGAAGTCAATC 212
Db |||||
QY 130 GCTAACTATAGCATTTATGCTCAGTATCAACATCTGAGCATTTTAAAGGCATATCAT 189
QY 213 CCGGCGTTGGTCCATACCCAGAAAAGGTCAGCCCATGCACTGCAAGAGTCATCTACA 272
Db |||||
QY 190 CTTTCTTGTGTCTATCTTACAAAGATGG---CGAGCGTGGCTTAAGCATTTCTACG 246
QY 273 CGCTTCGACTTCTTTTGGAGCTGAGCTTCAATTAATCTCAAAATCACTACTAGGTTGCA 332
Db |||||
QY 247 AGATACGGAACCCCGTTTGGAGCTAAGTCTTAATGCTCTGACTCCATAGTTCCGATACA 306
QY 333 TTCGAGCCCTCGGTTCCCTGACGGGACGAGGATGATCCATTCACACCCAGCGCATC 392
Db |||||
QY 307 TACGAGCCTATTAACGCGCAACTGGGAGCCATCTGGATCCGTTCAACACTTTTCGCTATC 366
QY 393 AGGCTGTCTTCCAGGACCTCAAGGCCATGTTTCCAGGGCTTGACCTGGAAATGTTTCGAT 452
Db |||||
QY 367 TGGGAGGCTCTAAAGAGCTTATAGATTCGCCAGCCAGGCATAGACCTTCAATGTTTCC 426
QY 453 CATTTCACTAAGCATTTGCTGTTTCGGAGGAAGAGGTCGGACTCTGCTAGATCGAGAT 512
Db |||||
QY 427 TACTTTAAACAGAGCTTACACTTTGACGCAACAGTCCACGTCACCTGCACTCTCAAAAC 486
QY 513 ATTGAGATCCCGCTCTTCAAGACACAGAACTGCGCAGCCGATCTGGAGCCATCTGGC 572
Db |||||
QY 487 TTGTCAGGAAACAGATCAAACTCAAAAGATAGGTTAGACCTTAA---AGGTGAC 543
QY 573 GATATGCTTTGAAGACTACATCTACCGCGGATCAAGTCTGATCGGAGCGGACGCCCA 632
Db |||||
QY 544 AAGTTGCTACTCAAGACCTACATCTACCCGAATTTGAAGTCCGTCGCAATCTGTTAAATCG 603
QY 633 AAAGAGAGACTCATGTTTGGCAAT---CAAGGCTCCGACACAGTTTGGCAAGTTGCC 689
Db |||||
QY 604 GTCAGGAGCTTGTGTTTGGCTCCGTCGCAAGCTAGCGCAGAGCATTAAGATATCCGT 663
QY 690 ACTCCACTGGCAATCCTCGAGGAGTTTAT--AGCTGAGCGAGACCCACCCCTCTCTCGG-- 745
Db |||||
QY 664 CTTGCTTTGAAATGCTAGAGACTATGTCAGTCTCGCAATAAAGTCCCTTACCCAGGAT 723
QY 746 -----CCACTTCTCTCATGCGATTTGCTCAAGCGCTCC 779
Db |||||
QY 724 GACAGTCAAACTACTCCATTAATCTTCAAGCCCTTCTCTTTCGACCTGAGTCTCTACC 783
QY 780 GAGTCCCGAATCAAGGCTTACTGTATGGAACGCCAGCTCGACCTGGCTCCATCGAAGT 839
Db |||||
QY 784 AAGTCTCGTCAAGATCTACTCTCTGGAACGAATGCTCTGTTGCCAGCGATGGAAGAT 843
QY 840 ATTTGAGACTCTCAACCGGCGACGGAAGATCCAGAGACACTGGATGTTGCTGAGTGGCTG 899
Db |||||
QY 844 CTTTGGACGCTTGGCGCGCTCGAGAAAGATCACTGTCACATTTGAGGGATTGGAGATGATC 903
QY 900 AGGAGGCTGTGCGAGCTATTTCCCGCTCAGGAGGCTGTGTCTCACTGCGCAACTGCTTT 959
Db |||||
QY 904 CGAGAACTTTGGGCTCTCTTAAATGCTCTCTGTTTGGCGCTTACCTGAGCCTTAC 963
QY 960 TAGAGCGGCTTACTCAACGAGAGAGCTCCCTTCAATTAATAATTTTACCTTGTCT 1019
Db |||||

Db 964 TTGCCCCCTGGCGCCATTCCTCAATAGCAACATTCCTGCTCCATGCGCCAAATACACCTTTACAC 1023
Qy 1020 CCTAAAGCCGCACTTCCGGAACACAGATCTATTTCCTGCTTTTGGCGAGAACGACAA 1079
Db 1024 CATATGATCCGATACAGAACCGCAAGTGTACTTTACTGTGTTCCGATGAATGATATG 1083
Qy 1080 ACCATCCGGAAGATGGCCACTTTTTCAGAGCAGAGGTTGGGGTGGCTTGGCTAAG 1139
Db 1084 GAGTGACTAATGCACTCACGAAATCTTTCATGAGGCATGAATGGAGCGATATGGCACT 1143
Qy 1140 AGCTATCAGCGGATTTGGCATCTTACTATCCGATGTGGACCTGAGACCGCAATAC 1199
Db 1144 AATACAAAGCCCTTGGGAACTTTCCCGCATATAATACGAAGCCCTTAATATAT 1203
Qy 1200 CTGAGCGGTGGATCTCTTCTTCAAGGGGAAAAACCGTACATGATGTGTACCTC 1259
Db 1204 ATCCACTCGTACATTTCTTCTCTACCGAAATAACAGCCATATTTAAGTGTGTATCTC 1263
Qy 1260 CATACCTTGAAGC 1273
Db 1264 CACTCATTTGAAC 1277

RESULT 4

US-09-518-657-6
; Sequence 6, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; APPLICANT: Wang, Jinghong
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; TITLE OF INVENTION: Which Encode Therefor and Related Methods
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; EARLIER FILING DATE: 2000-03-03
; EARLIER APPLICATION NUMBER: 60/125,490
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1598
; TYPE: DNA
; ORGANISM: Neotryphodium coenophialium
US-09-518-657-6

Query Match 12.7%; Score 176.6; DB 4; Length 1598;
Best Local Similarity 50.7%; Pred. No. 1.8e-47;
Matches 583; Conservative 0; Mismatches 529; Indels 39; Gaps 5;
Qy 33 GCAGCAACACTGTGCGCAAGCCCTTCTACGTGCTGAGTCAAGCCCTGAACTCTCGAAC 92
Db 34 GCAAGACACTCCACCAGGAAGTTTATCAACCCCTCAGCGAAACATTTTGACATTGCGAAC 93
Qy 93 AAGACACACAAATGGTGGTATAGACAGCTCCGATGTTTGGCCACCATGATGGGGG 152
Db 94 AATGACAGAGGTATGGTGGCACAGCGCGGCAATGTTCCAAAGATATCTCCAAACT 153
Qy 153 GCGGGTATGATGTTCCAGCACAGTCAAGTTCTCTGTATCCACCGTGAGGTCAATCATC 212
Db 154 GCTAATATAGCATTTATGCTAGTATCAACATCTGAGCATTTATAAAGCCATATCAT 213
Qy 213 CCGCGGTGGGTCCATACCCAGAAAAAGGCTAGCCCATGATGGAAGAGTCAATCTACA 272
Db 214 CTTTCTTGGTGTCTATCTCAAGAAAGTGG---CGAGCGGTGGCTAAGCATTTCTTAG 270
Qy 273 CGCTTCGGACTTCTTTCGAGTGAGCTTCAATTAATCTCCAAATCACTACTACGGTTTGA 332
Db 271 AGATACGAACCCGGTTTGGCTAAGTTTATGCTCTGACTCCATGTTCCGATATACA 330
Qy 333 TTGAGCCCTCGGTTCCCTTGACGGGAACGAAGATGATCCATTTCAACACCCAGGCAATC 392
Db 331 TAGGAGCCTATTAAACGCGCACTGGCAGCCACTTGGATCCGTTCAACACTTTTCGCTATC 390

Qy 393 AGGCTGTCTTCCAGGACCTCAAGGCCATGTTTCCAGGGCTTGACCTGGAAATGTTGAT 452
Db 391 TGGAGGCTCTTAAAAAGCTTATAGATTCAGCAGCCAGCATAGACCTTCAATGGTTTCC 450
Qy 453 CATTTCAATAAGATTTGGTGGTTTCGAGAGAGAGGCTCGGACTCTGCTAGATCGAGAT 512
Db 451 TACTTTAAACAAGAGCTTACACTTGACGCAACGAGTCCAGTACCTTGCACTCTTCAAAAC 510
Qy 513 ATTGAGATCCCGCTTTCAAGACACAGAACAACTGGCAGCCGATCTTGAGSCCATCTGGC 572
Db 511 TTGCTCAAGNAACAGATCAAACTCAAAACAAGCTAGCTTAGACCTTAA--AGGTGAC 567
Qy 573 GATATTGTTTGAAGACTACATCTACCGGGGATCAAGTCGATCGGACCGGAGACCCCA 632
Db 568 AAGTTCGTACTCAAGACCTACATCTACCCGAAATGAAGTCCGTCGCACTGGTAAATCG 627
Qy 633 AAAGAGAGACTCATGTTTGGACGAAT---CAAGGCTGCCGACAAAGTTTGGCAAGTTGCC 689
Db 628 GTCCAGGAGCTTGTTTGGTCCGTCGCAAGCTAGGCGAGAGCATAGAGTATCCGT 687
Qy 690 ACTCCACTGGCAATCTCGAGGAGTTTAT--AGCTGAGGAGCAGCCACCTCTCGG-- 745
Db 688 CTTGCTTTGAATGCTAGAAGACTATGTCAGTCTCGCAATAAAGTCCCTACACCGAT 747
Qy 746 -----CCACTTTCTCTCATGCGATTTTGGTCAAGCCGTCC 779
Db 748 GACAGTCAAAATCTCCATTTACTTACGGCTTCTCTTGGACCTGGTGGTCTCTTACC 807
Qy 780 GAGTCCCGAATCAAGGTCTACTGTATGGAACGCCAGCTCGACCTGGCCTCCATCGAAGT 839
Db 808 AAGTCTCGTCAAGATCTACTCTCTGGAAGATGCTCTCGTTGCCAGCATGGAAGAT 867
Qy 840 ATTTGGAATCTCAACGGCGAGAGCATCCAGAGACACTGGATGCTCTGGATGGCTG 899
Db 868 CTTTGGACGCTTGGCGCGCTCGAGAAAGATCAGTCCACTATTGAGGATTTGGAGATGATC 927
Qy 900 AGGAGCTGTGGCAGCTATTGCGCCGTCACGAGAGGCTGTGTCCACTGCGCGAACTGCTTT 959
Db 928 CGAGAACTTTGGGGTCTCTTAACTGTCTCTGTTTGGCGCTTACCCTAGCCTTAC 987
Qy 960 TACAGCGGGTACTCTCAGCGAGGAGCAGCTCCCTTCTATATAATTTTACCTTGTCT 1019
Db 988 TTGCCCCCTCGCGCCATTCGCAATGAGCAACTTCCGTCATGCGCAATTTACACCTTACAC 1047
Qy 1020 CTTAAAGCGCACTTCCCGAACCACAGATCTATTTCCTCTTTTGGGAGAGACGACAAA 1079
Db 1048 CATATGATCGATACAGAACCGCAAGTGTACTTTACTGTGTTGGCATGATGATG 1107
Qy 1080 ACCATCCGGAAGATTTGGCCACTTCTTTGAGAGCAGAGTTGGGTGGCTTGGCTAAG 1139
Db 1108 GAGTGACTAATGCACTCAGCAAAATTTCTCATGAGGCATGAATGGAGCGATATGGCAAGT 1167
Qy 1140 AGCTATCCAGC 1150
Db 1168 AATACAAAGC 1178

RESULT 5

US-09-518-657-5
; Sequence 5, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; APPLICANT: Wang, Jinghong
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; TITLE OF INVENTION: Which Encode Therefor and Related Methods
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; CURRENT FILING DATE: 2000-03-03
; EARLIER APPLICATION NUMBER: 60/125,490
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5
LENGTH: 1908
TYPE: DNA
ORGANISM: Neotyphodium coenophialum
US-09-518-657-5

Query Match 12.4%; Score 173.4; DB 4; Length 1908;
Best Local Similarity 50.5%; Pred. No. 2.3e-46;
Matches 581; Conservative 0; Mismatches 531; Indels 39; Gaps 5;

QY 33 GAGACACACTGTCGCAAGCCCTTCTAGTCTGAGTCAGCCCTGAACTCTCGAAC 92
DB 286 GCAAGACATCCACCAGGAAGTTTATCATACCTTAGGGAACATTTGACITTTGCAAC 345
QY 93 AAGGACACACAAATGGTGTATAGCAGAGTCGCGATTTGCGACCATGATGGGGG 152
DB 346 AATGACAGAGGCTATGGTGGACACACGCGCCCAATGTTGAAAAGATGCTCCAACT 405
QY 153 GCGGGTATGATGTTCAACGACAGTACAAGTTCTCTGTATCCACGCTGAGTCAATC 212
DB 406 GCTAACTATAGCATTTGATGCTAGTATCGACATCTGGCATTTATAAGAGCCATGCAIT 465
QY 213 CCGCGGTTGGTCCATACCAGAAAGGTCAGCCCATGCACTGGAAGATCATCTACA 272
DB 466 CTTTCTTGGTGTCTATCTCTACAGAGTGG---CGAGCGGTGGCTAAGCATTTCTACG 522
QY 273 CGCTTCGGACTTCCTTTGAGCTGAGCTCAATTAATCTCAAAATCACTACTAGGTTTGA 332
DB 523 AGATACGGAACCCGTTTGGAGCTAAGTCTTAATGCTCGGACTCGGTAGTTCGGTATACA 582
QY 333 TTCAGGCCCTCGGTTCCCTGACGGAACAGAGATGATCCATTCAACACCCAGGCAATC 392
DB 583 TACGAGCCTATTAAACCCGCACTGCGAGTCACTCTGGATCCGTTTAAACATTTTCGCTATC 642
QY 393 AGCCCTGTTCTCCAGGACCTCAAGGCCATGTTTCCAGGCTTCACTCGAATGTTTCGAT 452
DB 643 TGGAGGCCCTGAAAAGACATATTGATCCAGCCAGGATAGACCTTGAATGGTTTCT 702
QY 453 CATTCACATAAGCATTTGTCGTTTCGAGGAGAGGCTCGAGTCTGTAGATCGAGAT 512
DB 703 TACTTTAAACAGAGCTTACACTTGACGCAACAGAGTCCACGTACCTGCACTCGCAAAAC 762
QY 513 ATTGAGATCCCGTCTTCAAGACACAGAACTGGGAGCGGATCTGGAGCCATCTGGC 572
DB 763 TTGGTTAAGAACAGATCAAACTCAAAACAGCTCGCTTTGGACCTTAA---AGGTGAC 819
QY 573 GATATTGCTTTGAAGACCTACATCTACCGCGGATCAAGTCGATCGCAGCCGGGACCCCA 632
DB 820 AAGTTCGTACTGAAGACCTACATCTACCCGGAATTGAAGTCCGTCGCACTGTTAAATCG 879
QY 633 AAGAGAGACTCATGTTTGAAGCAAT---CAAGGCTGCGGACAAAGTTTGGCAAGTTGCC 689
DB 880 GTCCAGGAGCTCGTGTGTTGGCTCCGTCGCAAGCTAGCGCAGAAAGCACAAAGATATCCGT 939
QY 690 ACTCCACTGGCAATCCTCGAGAGTTTAT--AGCTGAGGAGCACCACCCCTCTCTCGG-- 745
DB 940 CTTGCTTTGAAATGCTAAGACTATGTCAGTCTCGCAATAAATCTCTACCAAGGAT 999
QY 746 -----CCACTTCTCTCATGCAATTTGGTCAAGCGGTCC 779
DB 1000 GACAGTCACAATACTTATTATCTTCAACGCTTCTCTTGGACCTGTAAAGTCTCTACC 1059
QY 780 GAGTCCCGAATCAAGCTCTACTGTATGGAACCGCAGCTCGACTGCGCTCCATCGAAGT 839
DB 1060 AAGTCTCGTGTCAAGATCTACTCTCTGGAAACGAATGGTCTCGTTCAGCGATGGAAGAT 1119
QY 840 ATTGGACTCTCAACGGGGGAGGAGACGATCCAGAGACACTGGATGGTCTGGATCGGTG 899
DB 1120 CTTTGGACGCTTGGCGGCGCTCGAGAAGATCAGTCCACTATTGAGGATTTGAGATGATC 1179
QY 900 AGGAGCTGTGGCAGATATGCCCTGACGAGGAGTCTGTCTCACTGCGCAACTGCTTT 959
DB 1180 CGAAGACTTTGGGGTCTCTCAACATGCTCTCTGTTTGGCGGCTTACCTGAGCCTTAC 1239

QY 960 TACGAGCGGGTACTCTACCCGAGGAGAGCTCCCTTCAATATAAATTTTACCTGTCT 1019
DB 1240 TTGCCCTTCGGCGGCAATTCOAATGAGCAATTCGCTCCATGAGCACTTACAC 1299
QY 1020 CCTAAAGCGCACTTCCGAAACCAAGATCTATTTCCTGCTTTTGGGCAAGACACAAA 1079
DB 1300 CATATGATCCATACAGAACCCGCAAGTGTACTTTACTGTGTTCCGATGAATGATG 1359
QY 1080 ACCATCGCGGAAGATTGGCCACCTTTCTTGAAGCAGAGGTTGGGTGGCTTGGCTAAG 1139
DB 1360 GAGGTGATTAATGACTCAGCAGCATTTCTTATGAGGATGAATGAGCGGATATGCGCAAGT 1419
QY 1140 AGCTATCCAGC 1150
DB 1420 AATACAAAGC 1430

RESULT 6

US-09-010-928B-1
Sequence 1: Application US/09010928B
Patent No. 5994099
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
TITLE OF INVENTION: CODING THEREFOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATEHOUSE RD. SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: -
LOCATION: 1..2830 /note= "Flagelliform DNA sequence
OTHER INFORMATION: taken from the 5' region. The putative start codon is at
OTHER INFORMATION: position 219"
FEATURE:
NAME/KEY: CDS
LOCATION: 219..2830
US-09-010-928B-1

Query Match 2.7%; Score 37.8; DB 2; Length 2830;
Best Local Similarity 52.9%; Pred. No. 0.11; 72; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 0; Gaps 0;
QY 435 GACCTGAATGGTTCGATCATTTCACTAAAGCATTCGTCGTTGCGAGGAGGCTCGG 494

Db 1906 GACCTGGTAGCTTTGGACAGGAGCGTAGGACCTGGTCCGACCGAGGCGGTAG 1965
Qy 495 ACTCTGCTAGATCGAATATTGAGATCCCGCTCTTCAAGACACAGAACTGGAGCC 554
Db 1966 GATCTGCTGGCTCCGACAGGAGGAGTAAGACCTAGTGGCTCCGACCGAGTGGCGTAG 2025
Qy 555 GATCTGAGCACTCTCGGATATTGCTTGAAG 587
Db 2026 GAATGAGCGCTAGACCGGTGGTGGAG 2058

RESULT 7
US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match 2.5%; Score 35.2; DB 4; Length 399;
Best Local Similarity 13.4%; Pred. No. 0.24;
Matches 49; Conservative 147; Mismatches 170; Indels 0; Gaps 0;
Qy 802 GTATGGAAGCCAGCTCGACCTGGCTCCATCGAAGGTATTGGACTCTCAACGGCGGAC 861
Db 377 GTATGAGATCAATCAGCAAGKKGKGGTGMKMTWYKYSYMSKMTWMTYAK 138
Qy 862 GGAACATCCAGACACTGGATGCTGCTGATCGCTGAGGAGCTGTGGCAGCTATTGC 921
Db 317 MGRAKMKRKTCTKTKYMWKGGKGGSTYMAESRRGSGTGRWSYRAMWRGSKSWGGSY 258
Qy 922 CGTACGAGGAGTCTGTGCTCCTGCGCACTGCTTTTACGAGCGGGTACTCACCGC 981
Db 257 YRMAGYRSSRWSYSAWRRKMTCKWGRSSWSRSTGYAWMYKKSMTSRKMYKK 198
Qy 982 AGGAGAGCTCCCTTCATATATAATTTTACCTTGCTCTCTAAAGCGCACTCCCGAAC 1041
Db 197 RRKWRKCTSTRTCYRGSTYKWKAYTKKRWTRWYKYSYMSKMTWMTYAK 138
Qy 1042 CACAGATCTATTCCTGCTTTTGGCAGACAGCAAAACCATCCGCGGAAGATTGGCA 1101
Db 137 YWTKRWKMTSTKWTCTWCTWCKCTTYWMAGMTMYRYRYWYAKRWKSKRTWSTTCYOM 78
Qy 1102 CTTCTTTGAGACAGAGGTGGGTGGCTTGCTTAAGAGCTATCCAGCGGATTGGCAT 1161
Db 77 KYMAKCVSWNSMMSMKGKSMWIKWTYYYYYMMKSKMTYWSMSYARKCWRTYAK 18
Qy 1162 CCTACT 1167
Db 17 TYTWT 12

RESULT 8
US-09-252-991A-10059/c
; Sequence 10059, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10059
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10059
Query Match 2.5%; Score 34.8; DB 4; Length 606;
Best Local Similarity 50.0%; Pred. No. 0.41;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 508 GAGATATTGATCCCGCTCTTCAAGACACAGAACTGGCAGCGGATCTGGAGCCAT 567
Db 466 GCGGAGTGACATCAGCGCTGTCGAGAGCCAGAACAGCCCAATATCGGCTCGGGCGTCA 407
Qy 568 CTGGCGATATTGTTTGAAGACCTACATCTACCCCGGGATCAAAGTCGATCGCGACCGGGA 627
Db 406 ATGCGTGACTGTGTCGAGAGCCAGACCGTGCCCAATATCGGTTCTGGCGCAAGTGGA 347
Qy 628 CCCAAAAGAGAGACTCATGTTTGAAGCAATCAAGGTCGCGCAAGTTTGGCA 681
Db 346 TCACCGTTGTCAAAGCCAGAAATGCGGCGCAATATCGGTTTCAGGCGGAGTGGA 293

RESULT 9
US-09-252-991A-9867
; Sequence 9867, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9867
; LENGTH: 4248
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9867

Query Match 2.5%; Score 34.8; DB 4; Length 4248;
Best Local Similarity 50.0%; Pred. No. 1.4;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 508 GAGATATTGATCCCGCTCTTCAAGACACAGAACTGGCAGCGGATCTGGAGCCAT 567
Db 1367 GCGGAGTGACATCAGCGCTGTCGAGAGCCAGAACAGCCCAATATCGGCTCGGGCGTCA 1426
Qy 568 CTGGCGATATTGTTTGAAGACCTACATCTACCCCGGGATCAAAGTCGATCGCGACCGGGA 627
Db 1427 ATGCGTGACTGTGTCGAGAGCCAGAACCGTGCCCAATATCGGTTCTGGCGCAAGTGGA 1486
Qy 628 CCCAAAAGAGAGACTCATGTTTGAAGCAATCAAGGTCGCGCAAGTTTGGCA 681
Db 1487 TCACCGTTGTCAAAGCCAGAAATGCGGCGCAATATCGGTTTCAGGCGGAGTGGA 1540

RESULT 10
US-07-669-171-1
; Sequence 1, Application US/07669171
; Patent No. 5304541

GENERAL INFORMATION:
APPLICANT: PURCHIO, ANTHONY F.
APPLICANT: MADISEN, LINDA
APPLICANT: MEWIN, JUNE RAE
TITLE OF INVENTION: TGF-b1/2: A NOVEL CHIMERIC TRANSFORMING
GROWTH FACTOR-BETA
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/07/669,171
FILING DATE: 19910314
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-159-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1560 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 261..1430
US-07-669-171-1

Query Match 2.4%; Score 33.6; DB 1; Length 1560;
Best Local Similarity 48.4%; Pred. No. 1.8;
Matches 93; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 812 CGAGCTGCAGCTGGCTCCATCGAAGGTATTGGACTCTCAACGGGCGACGGAACGATCC 871
DB 1217 CAACCTCTGCTGGGCGCTGTCCCTACATTTGGAGCTGGACAGCAGTACAGCAAGT 1276
QY 872 AGAGACACTGGTGTCTGGATCGCTGAGGAGCTGTGCAGCTATTGCCCGTCAACGA 931
DB 1277 CTGGCCCTGTCAACACAGATTAACCCGGCGCTCGGCGGCGCGTGTGCTGCGCGCA 1336
QY 932 GGCTCTGTCTCCACTGCCGAACCTTTTACGAGCGGGGTACCTCACCGCAGGAGCAGCT 991
DB 1337 GGGCGCTGGAGCCACTGCCCATCTGTACTAGTGGCGCGCAAGCCCAAGGTGGAGCAGCT 1396
QY 992 CCCCTTCATTAT 1003
DB 1397 GTCCAAACATGAT 1408

RESULT 11
US-09-358-082A-28
Sequence 28, Application US/09358082A
Patent No. 6589784
GENERAL INFORMATION:
APPLICANT: Antoniou, Michael
APPLICANT: Crombie, Robert
TITLE OF INVENTION: A Polynucleotide
FILE REFERENCE: CAC00056

CURRENT APPLICATION NUMBER: US/09/358,082A
CURRENT FILING DATE: 1999-07-21
PRIOR APPLICATION NUMBER: GB 9815879.3
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: US 60/107688
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: GB 9906712.6
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/127410
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: GB 9909494.8
PRIOR FILING DATE: 1999-04-23
PRIOR APPLICATION NUMBER: US 60/134016
PRIOR FILING DATE: 1999-05-12
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 28
LENGTH: 9098
TYPE: DNA
ORGANISM: Homo sapiens
US-09-358-082A-28

Query Match 2.4%; Score 32.8; DB 4; Length 9098;
Best Local Similarity 55.2%; Pred. No. 9.8;
Matches 64; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 1278 AGTCTGCTGCCAGGAGGTGGCTATGTGTACGATGGCCCAATCCTTAGGACTAGTTT 1337
DB 13 AGTCAGCCCGACAGGAGCTGGGATGAGGATATATACAGGCATTCTGTTAATGCTGATT 72
QY 1338 ATCCCTTCATTCTATGATCGTTGAATGTGGTCCGAAAAAAGAAAAA 1393
DB 73 GTTCTTATCTCTATCTATATCTGACGTGTTTCACAAAAAAGAAAAA 128

RESULT 12
US-09-103-840A-2/c
Sequence 2; Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.4%; Score 32.8; DB 3; Length 4403765;
Best Local Similarity 49.4%; Pred. No. 1.2e+02;
Matches 85; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 699 GCAATCTCGAGGAGTTTATAGCTGAGGAGCAGCCACCTCCTCGGCCACTTTCTCTCA 758
DB 291460 GCAACCATCGCCGAGGTGTTGCTCGGACGGCGCCACGTTGTCGCATCGATGTGAG 291401
QY 759 TGCATTGTTCAAGCCGCTCCGAGTCCCGAATCAAGTCTACTGTATGGAACGCCAGCTC 818
DB 291400 TCCGCGCCGAGAACCTGGCCGAAACCCAGCAGCAAGGTGGCGCGCACCGCGCTGTGCTC 291341


```

; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TcB-se
US-09-168-595-60

```

Query Match	2.3%; Score 32.6; DB 4; Length 2416;
Best Local Similarity	47.0%; Pred. No. 5.1;
Matches 101; Conservative 0; Mismatches 114; Indels 0; Gaps 0;	
QY	39 ACACGTGTCACAAAGCCCTTCTACGTGCTGAGTCAAGCCCTGAACCTCTCGAACCAAGAC 98
DB	152 ACTCACCGCCCAAGGCCCTTCAGCCGCCATGCACCTCTGATGCAGTCAGAGAGTA 211
QY	99 CACACAAATGGTGGTATAGCACAGCTCCGATGTTGGCCACCATGATGCGGGGCGCGC 158
DB	212 GAGGACAGTGGCGGGCAGCGCGATCTGTGATTGGGTCTGGCCCTCAGCTACTACCCAC 271
QY	159 TATGATGTTACGGCACAGTACAAGTTCTCTGTATCCACGTTGAGTCAATCATCCCGCG 218
DB	272 TACTATGATGATGCCCGGACCATGTACAGGTGTTCCGCCGTGGCTTAGCATCTCAGG 331
QY	219 TTGGTTCATACCCAGAAAGGTCAGCCCATGCA 253
DB	332 AATGGGCCCTGCTCTTGGTTTCAGGAAGCCCTAAGCA 366

Search completed: July 29, 2004, 01:19:52
Job time : 143 secs

OM nucleic - nucleic search, using sw model

Run on: July 29, 2004, 00:10:36 ; Search time 687 Seconds
(without alignments)
9941.884 Million cell updates/sec

Title: US-10-099-704-1
Perfect score: 1393
Sequence: 1 ccgaagtgcaatgag.....cgaaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:**
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1393	100.0	1393	14	US-10-099-704-1
2	58.2	4.2	992	16	Sequence 1, Appli
3	48.2	3.5	749	16	Sequence 36657, A
4	43.6	3.1	1292	16	Sequence 36210, A
5	38.6	2.8	1184	13	Sequence 37019, A
6	38.6	2.8	1184	15	Sequence 394, App
7	38.6	2.8	1184	15	Sequence 394, App
8	38.6	2.8	1184	15	Sequence 394, App
9	38.6	2.8	1184	15	Sequence 394, App
10	38.6	2.8	1184	15	Sequence 394, App
11	38.6	2.8	1184	15	Sequence 394, App
12	38.6	2.8	1184	16	Sequence 394, App
13	38.6	2.8	1184	16	Sequence 394, App
14	38.6	2.8	1184	16	Sequence 394, App

C 15	38.6	2.8	1184	16	US-10-141-759-394	Sequence 394, App
C 16	38.6	2.8	1184	16	US-10-140-803-394	Sequence 394, App
C 17	38.6	2.8	1184	16	US-10-140-864-394	Sequence 394, App
C 18	37.2	2.7	2289	17	US-10-437-963-71390	Sequence 71390, A
C 19	37.2	2.7	2686	17	US-10-437-963-71401	Sequence 71401, A
C 20	36.4	2.6	66499	17	US-10-322-281-39	Sequence 39, Appl
C 21	35.8	2.6	13288	13	US-10-332-288-27	Sequence 27, Appl
C 22	35.6	2.6	1258	13	US-10-282-122A-19400	Sequence 19400, A
C 23	35.4	2.5	945	16	US-10-260-238-4003	Sequence 4003, Ap
C 24	34.8	2.5	3237	9	US-09-727-205-1	Sequence 1, Appli
C 25	34.6	2.5	4254	13	US-10-282-122A-30494	Sequence 30494, A
C 26	34.6	2.5	1995	16	US-10-369-493-37566	Sequence 37566, A
C 27	34.4	2.5	1065	13	US-10-451-207-10	Sequence 10, Appl
C 28	34.4	2.5	2454	15	US-10-156-761-3622	Sequence 3622, Ap
C 29	34.4	2.5	9025608	15	US-10-156-761-1	Sequence 1, Appli
C 30	34.2	2.5	571	13	US-10-142-426-542	Sequence 542, App
C 31	34.2	2.5	571	15	US-10-123-155-542	Sequence 542, App
C 32	34.2	2.5	571	15	US-10-146-731-542	Sequence 542, App
C 33	34.2	2.5	571	15	US-10-140-473-542	Sequence 542, App
C 34	34.2	2.5	571	15	US-10-141-761-542	Sequence 542, App
C 35	34.2	2.5	571	15	US-10-142-885-542	Sequence 542, App
C 36	34.2	2.5	571	15	US-10-158-790-542	Sequence 542, App
C 37	34.2	2.5	571	16	US-10-137-871-542	Sequence 542, App
C 38	34.2	2.5	571	16	US-10-140-923-542	Sequence 542, App
C 39	34.2	2.5	571	16	US-10-141-756-542	Sequence 542, App
C 40	34.2	2.5	571	16	US-10-141-759-542	Sequence 542, App
C 41	34.2	2.5	571	16	US-10-140-805-542	Sequence 542, App
C 42	34.2	2.5	571	16	US-10-140-864-542	Sequence 542, App
C 43	34.2	2.5	624	13	US-10-027-632-189298	Sequence 189298,
C 44	34.2	2.5	624	16	US-10-027-632-189298	Sequence 189298,
C 45	34.2	2.5	4674	17	US-10-437-963-69259	Sequence 69259, A

ALIGNMENTS

RESULT 1
US-10-099-704-1
; Sequence 1, Application US/10099704
; Publication No. US20020197682A1
; GENERAL INFORMATION:
; APPLICANT: Christensen, Bjorn Eggert
; APPLICANT: Mollgaard, Henrik
; APPLICANT: Kaasgaard, Svend
; APPLICANT: Lehbeck, Jan
; TITLE OF INVENTION: Methods for producing polypeptides in
; FILE REFERENCE: 4483.200-US
; CURRENT APPLICATION NUMBER: US/10/099,704
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/472,364
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/139,593
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/117,396
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: PA 1999 00745
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: PA 1998 01726
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer 5956
; NAME/KEY: CDS
; LOCATION: (15)...(1328)
US-10-099-704-1

Query Match 100.0%; Score 1393; DB 14; Length 1393;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAACGACCAACACTGCTGCCAAGCCCTTCT 60
DB 1 CCGAAGCTGAGCAATGGAGATCTCCAGAAACGACCAACACTGCTGCCAAGCCCTTCT 60

QY 61 ACGTGCTGAGTCAAGCCCTGAACTCTCGAACAAGGACCAACAAAATGGTGTATAGCA 120
DB 61 ACGTGCTGAGTCAAGCCCTGAACTCTCGAACAAGGACCAACAAAATGGTGTATAGCA 120

QY 121 CAGTCCGATGTTGCCACCATGATGGCGGGGCGGCTATGATGTTCCAGCACAGTACA 180
DB 121 CAGTCCGATGTTGCCACCATGATGGCGGGGCGGCTATGATGTTCCAGCACAGTACA 180

QY 181 AGTTCCTCTGTATCCACCGTGAGGTGATCATCCCGGCTTGGGTCCATACCCAGAAAGG 240
DB 181 AGTTCCTCTGTATCCACCGTGAGGTGATCATCCCGGCTTGGGTCCATACCCAGAAAGG 240

QY 241 GTGAGCCCATGCACTGGAAGATCATCTCAAGCTTCGACATTCCTTTGCGAGCTGAGCT 300
DB 241 GTGAGCCCATGCACTGGAAGATCATCTCAAGCTTCGACATTCCTTTGCGAGCTGAGCT 300

QY 301 TCAATTACTCCAAATCACTACTACGTTTGCATTCGAGCCCTCGGTTCCCTGACGGGAA 360
DB 301 TCAATTACTCCAAATCACTACTACGTTTGCATTCGAGCCCTCGGTTCCCTGACGGGAA 360

QY 361 CGAAGGATGATCCATCAACCCAGCAATCAGGCTGTCTTCCAGGACCTCAAGGCCA 420
DB 361 CGAAGGATGATCCATCAACCCAGCAATCAGGCTGTCTTCCAGGACCTCAAGGCCA 420

QY 421 TGGTTCAGGCTTGACCTCGAATGTTGATCATTTCAATTAAGCATTTGGTTCGTTTGG 480
DB 421 TGGTTCAGGCTTGACCTCGAATGTTGATCATTTCAATTAAGCATTTGGTTCGTTTGG 480

QY 481 AGAAGAGGCTCGGACTCTGCTAGATCGAGATATTCAGATCCCGCTTTCACAGACACAGA 540
DB 481 AGAAGAGGCTCGGACTCTGCTAGATCGAGATATTCAGATCCCGCTTTCACAGACACAGA 540

QY 541 ACAAAGTGGAGCGGATCTGGAGCCATCTGGGATATTTCTTGAAGACCTACATCTACC 600
DB 541 ACAAAGTGGAGCGGATCTGGAGCCATCTGGGATATTTCTTGAAGACCTACATCTACC 600

QY 601 CGCGGATCAAGTCAAGTTCGGACCCCGGACCCCAAGAGAGATCTATGTTTGAAGCAATCA 660
DB 601 CGCGGATCAAGTCAAGTTCGGACCCCGGACCCCAAGAGAGATCTATGTTTGAAGCAATCA 660

QY 661 AGCTCCGACAAAGTTTGGCAAGTTGCGCATCTCCACTGCGCAATTCCTCGAGGAGTTATAG 720
DB 661 AGCTCCGACAAAGTTTGGCAAGTTGCGCATCTCCACTGCGCAATTCCTCGAGGAGTTATAG 720

QY 721 CTGAGCGAGCACCCACCTCTCGGCGCATTTCTCTCATGCGATTTGGTCAAGCGCTCG 780
DB 721 CTGAGCGAGCACCCACCTCTCGGCGCATTTCTCTCATGCGATTTGGTCAAGCGCTCG 780

QY 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGCAGCTCGACTGCGCTCCATCGAAGGTA 840
DB 781 AGTCCCGAATCAAGGTCTACTGTATGGAACGCGCAGCTCGACTGCGCTCCATCGAAGGTA 840

QY 841 TTTGGACTCTCAAGCGGCGGACGATCCAGAGACCTGATGGTCTGGATGGCTGCA 900
DB 841 TTTGGACTCTCAAGCGGCGGACGATCCAGAGACCTGATGGTCTGGATGGCTGCA 900

QY 901 GGGAGCTGTGGAGCTATTGCCCGTCAAGAGGGTCTGTGTCCATGCGGAACTCTTTT 960
DB 901 GGGAGCTGTGGAGCTATTGCCCGTCAAGAGGGTCTGTGTCCATGCGGAACTCTTTT 960

QY 961 ACGAGCGGGTACCTCAAGCGAGGAGGAGCTCCCTTCATTAATTTACCTTGTCTC 1020
DB 961 ACGAGCGGGTACCTCAAGCGAGGAGGAGCTCCCTTCATTAATTTACCTTGTCTC 1020

QY 1021 CTAAAGCGCACTTCCCGAACACAGATCTATTTCCTGCTTTTGGGAGAACACGACAAA 1080

DB 1021 CTAAAGCGCACTTCCCGAACACAGATCTATTTCCTGCTTTTGGGAGAACGACAAA 1080

QY 1081 CCATCGCGGAGGATTTGGCCACCTTTTGGAGACAGAGGTTGGGTGGCTTGGCTAAGA 1140
DB 1081 CCATCGCGGAGGATTTGGCCACCTTTTGGAGACAGAGGTTGGGTGGCTTGGCTAAGA 1140

QY 1141 GCTATCCAGCGGATTTGGCATCTTACTATCCCGATGTGGACCTGCAGACCGCAATCACC 1200
DB 1141 GCTATCCAGCGGATTTGGCATCTTACTATCCCGATGTGGACCTGCAGACCGCAATCACC 1200

QY 1201 TGCAGGCGTGTATCTCTTCTTACAAGGGGAAAAACCGTACATGATGTGTACTCTC 1260
DB 1201 TGCAGGCGTGTATCTCTTCTTACAAGGGGAAAAACCGTACATGATGTGTACTCTC 1260

QY 1261 ATACTTCGAAAGCGTTCAAGTCTGCTGCCAGGAGGTGGCTATGTGTCAAGATGGCCACA 1320
DB 1261 ATACTTCGAAAGCGTTCAAGTCTGCTGCCAGGAGGTGGCTATGTGTCAAGATGGCCACA 1320

QY 1321 ATCTTAGGACTAGTTTATCCCTTCAATTCATGATCCGTTGAATGTGTGTGCGAAAAA 1380
DB 1321 ATCTTAGGACTAGTTTATCCCTTCAATTCATGATCCGTTGAATGTGTGTGCGAAAAA 1380

QY 1381 AAAAAA 1393
DB 1381 AAAAAA 1393

RESULT 2
US-10-369-493-36657
; Sequence 36657, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 36657
; LENGTH: 992
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-36657

Query Match 4.2%; Score 58.2; DB 16; Length 992;
Best Local Similarity 48.2%; Pred. No. 1.5e-08;
Matches 259; Conservative 0; Mismatches 268; Indels 10; Gaps 3;

QY 429 GGGCTTGACCTGGAATGGTTTCGATCATTTTCACTAAAGCATTTGGTTCGTTTCGAGGAGAG 488
DB 403 GGATTCGATATGACATCTTCGATCATGTTCCGGATACATCTTCAATTTTCGACGAGGAG 462

QY 489 GCTCGGACTCTGTAGATCGAGATATTGATATCCCGTCTTCAAGACACAGAACAACTG 548
DB 463 ACGGCTGCCCTTCAAAACCCAGAT-----GCCTATGGGATCAAGACCGATATCTGCTCTA 516

QY 549 GCAGCCGATCTGAGCCATCTGGCGATATTGTCTTGAAGACCTTACATCTACCGCGGATC 608
DB 517 GGATTCGATATGAG---GGGACACACGTTGGTCAAGTGTATATGATCTCGCGGTGG 573

QY 609 AGTTCGATCGGACCGGGGACCCCAAGAGAGACTCATGTTTGAAGCAATCAAGGCTGCC 668
DB 574 AAGAGCTTTGCAACGGGGGTTCCGGTTGCAAGCTGATACGAGATTCGCTTTCGAGCGGATC 633

QY 669 GACAAGTTTGGCAAGTTCGCACTCCACTGGCAATCTTCGAGGAGTTTATAGCTGAGCGA 728

Db 634 AAGACACAGTTCGATTGTAACAGCGCTCGAACTGGTTGATGATATATGAGGATAGT 693
QY 729 GCACCCACCTCTCTCGGCACCTTTCTCTCATGCGATTTCGTCAGGCGTCGAGTCCCGA 788
Db 694 GGTCTGGGATTTAAGGAGGTTTCATGCTGGGACTGTGTGCTTTGGCCCGAGCGGA 753
QY 789 ATCAAGGTCTACTGTATGGAACGCCAGCTCGACCTTGCCCTCCOATCGAAGTATTTGGACT 848
Db 754 CTCAAAATTTACGGCATCATCAACGAGGTCTCGCTGGGCAAAAGTTGAGGAACGTGTGACG 813
QY 849 CTCACGCGGCGGAGCAACCATCCAGAGACACTGGATGCTCTGGATGCGCTGAGGGAGCTG 908
Db 814 ATGGGTGGCGGCTCAACGACGAGCACTCTCGAGGACTCTC-GCTCTTCGGCGCT 872
QY 909 TGGACCTATTGCCCTCGAGGAGTCTGTGTCTCACTCCGAACTGCTTTTACGAG 965
Db 873 TTGCACCTACTGGAGGTCAATAGGATGACCGGCTGTTTTCGAAGGGGAGGAG 929

RESULT 3
US-10-369-493-36210
; Sequence 36210, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 36210
; LENGTH: 749
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-36210

Query Match 3.5%; Score 48.2; DB 16; Length 749;
Best Local Similarity 45.3%; Pred. No. 3.6e-05;
Matches 341; Conservative 0; Mismatches 398; Indels 14; Gaps 4;

QY 57 TTCTACGTGCTGAGTCAAGCCCTGAACCTCTGGAACAGAGCACACAAAATGGTGGTAT 116
Db 1 TTTGACGTGCTGACCTCAGCCTTGTCTTTTCCACCAGAGACGAGCAATGGTGGCGC 60

QY 117 AGCACAGCTCCGATGTTGCCACCATGATGGGGGGCGGCTATGATGTTTCACGCACAG 176
Db 61 AGACCGGCCAAATGTTGGTCAGATGCTCGCTCGCTGTGGCTATACCTCGATCAGCAG 120

QY 177 TACAAGTTCTCTGTATCCACCGTGAGTCAATCCCGCGTGGGTCCATACCAGAA 236
Db 121 TATCGCACCTACCTTCTACTACCAACCACTGTTCCCGCGCTCGGCCCTCAGCCAGCA 180

QY 237 AAGGTTCAGCCATGACCTGGAAGATCATCTCAGAGCTTCGGACTTCCTTTGAGCTG 296
Db 181 ACATTCCATCCAGTCTGACTGTACGCGGTT---ACCATGGAGTTCAGCTCAACTAC 237

QY 297 AGCTTCAATTAATCCAAATCACTACTACGTTTGTGATTTCGAGCCCTCGGTTCCCTGAGC 356
Db 238 CAGCAAAAGGTGCGCATCAATGGTCCGATTTGGCGGGAACCTATCGACTCCCTTTCG 297

QY 357 GGAAACAGGATGATCCATTCAACAC--CAGCAATAGGCTGTCTTCCAGGACCTCA 414
Db 298 GGGACGGAACGGGACCCATTTATATCAGATCCCGCGCGGAGATGGTAAACACTTCTCCA 357

QY 415 AGSCCATGTTCCAGGGCTTGACCTGGAATGGTTCGATCATTTCACTAAAGCATTTGGTGC 474

Db 358 GAGCGGAGTTAAAGGATTCGATCCGGAGCTTTATGCTACTTTTATGCCAAAGCATTTCTC 417
QY 475 TTTTCGAGGAGAGGCTCGGACTCTGTAGATCGAGATATTAGATCCCGTCTTCAAGA 534
Db 418 TAACTCTGTGAGCAGCAAGCCAGACTACGGAAGAGTACTCTGGTGTGACAAGTTAAGA 477
QY 535 CACAGAACAACTGGCAGCCGATCTGGAGCCATCTGGCGATATTTGTCTTGAAGACCTTACA 594
Db 478 CGCAATATGCTTTTCGGGTTCGATTTTAAGGG---TGATGAGGTTTCACTGAAGGGGTATA 534
QY 595 TCTACCGCGGATCAAGTTCGATCGGACCGGAGCCCAAAAGAGAGACTCATGTTTTCAGC 654
Db 535 GCTATCCCGGCTGAAAGCCACAAATGGCAGCCAGGAAGTTGGAAGCTGTTCGAGAGACG 594
QY 655 CAATCAAGGCTGCCGACAAAGTTTGGCAAGTTG-----CCACTTCCACTGGCAATCTCTG 708
Db 595 GGGTCAAGGACCTGAAAAAACCAGGCAAACTGGACTGCACCGAGGCTCTGGCAGCTGTGG 654
QY 709 AGGAGTTTATAGCTGAGGAGCACCACCTCTCTCGGGCCACTTTCTCTCATGCGATTGG 768
Db 655 AAGCTTACATGACTGAACCTCAACAACTGGGGCTACCAACACCTCTGGGCATGGGATTACG 714
QY 769 TCAAGCCGTCCGAGTCCCGAATCAAGGTCTACT 801
Db 715 TCAAGCCGTCCGAATCAGCTCTCAAGTTGTATT 747

RESULT 4
US-10-369-493-37019
; Sequence 37019, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37019
; LENGTH: 1292
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-37019

Query Match 3.1%; Score 43.6; DB 16; Length 1292;
Best Local Similarity 50.0%; Pred. No. 0.0018;
Matches 109; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 269 CACACGCTTCGGACTTCCTTTTCGAGCTGAGCTTCAATTACTTCCAAATCACTACTACGGTT 328
Db 294 CACGCCCTTCGAATTCATGGAATGGAACAATCCCGCGGCGAGCCAGATGTCGCTT 353

QY 329 TGCATTCGAGCCCTCGGTTCCCTGACGGGAACGAAGATGATCATCAACCCAGGC 388
Db 354 CGGTTTACGCTATAGCAACCAATGGCAGGAACGTCCTTAGATCCCTTTGAACCACTTGGC 413

QY 389 AATCAGGCGCTTCTCTCAGGACCTCAAGGCGCATGGTTCCAGGGCTTGACCTGGAATGGTT 448
Db 414 CATGCGGAGATCTCTACAAGCTCTCCTCGGCTGTCCAGGCTCTCATCTCAGCTGAC 473

QY 449 CGATCATTTCTAAAGCATTTGCTCGTTTCGGAGGAAG 486
Db 474 ACATCATTTCTCGCCACTCTGTTTCGATCATGACTATG 511

RESULT 5

```

US-10-142-426-394/c
; Sequence 394, Application US/10142426
; Publication No. US20040048333A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C224
; CURRENT APPLICATION NUMBER: US/10/142,426
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-426-394

```

```

Query Match      2.8%; Score 38.6; DB 13; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY 92 CAAGGACACACAAAATGGTGTATAGCACAGCTCCGATGTTGGCCACCATGATGGCGG 151
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 GCGCGGTATGATGTTCCACGACAGTACAAAGTTCCTCTGTATCCACCGTGAGGTCA 211
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 930 .Y.CS.M..SBMCN.DY.YY.ATSS.T..Y.T..SC.HCT.YCA.SCTBM..YSY..CAT 871
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 CCGCGGTGGTCCATACCCAGAAAGGTGAGCCCATGCGGAGGATGATCTCATCTCAC 271
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 870 .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACCTTCGGACTTCCTTCGAGCTGAGCTTCAATTAATCACTCAATCACTACTACGTTTC 331
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 810 .HWCT.KKT.MBHMD.S.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS..T.YGN 751
QY 332 ATTGAGCCCTCGGTTCCCTGAGCGGAGGAGGATGATCCATTCAACACCCAG 386
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 750 RTYNHM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

```

```

RESULT 6
US-10-123-155-394/c
; Sequence 394, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.

```

```

; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-394

Query Match      2.8%; Score 38.6; DB 15; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY 92 CAAGGACACACAAAATGGTGTATAGCACAGCTCCGATGTTGGCCACCATGATGGCGG 151
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 GCGCGGTATGATGTTCCACGACAGTACAAAGTTCCTCTGTATCCACCGTGAGGTCA 211
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 930 .Y.CS.M..SBMCN.DY.YY.ATSS.T..Y.T..SC.HCT.YCA.SCTBM..YSY..CAT 871
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 CCGCGGTGGTCCATACCCAGAAAGGTGAGCCCATGCGGAGGATGATCTCATCTCAC 271
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 870 .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACCTTCGGACTTCCTTCGAGCTGAGCTTCAATTAATCACTCAATCACTACTACGTTTC 331
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 810 .HWCT.KKT.MBHMD.S.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS..T.YGN 751
QY 332 ATTGAGCCCTCGGTTCCCTGAGCGGAGGAGGATGATCCATTCAACACCCAG 386
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 750 RTYNHM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

```

```

RESULT 7
US-10-146-731-394/c
; Sequence 394, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C323
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15

```

```

; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-146-731-394

Query Match
Best Local Similarity 2.8%; Score 38.6; DB 15; Length 1184;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY 92 CAAGGACCACAAAATGGTGTATAGCACAGCTCCGATGTTTGCACCATGATGCGGG 151
Db 990 C.MT.RMNCYDN.M.....C.D.S...S...B.SHA...B.NY.T.TTBS.YB.SY.. 931
QY 152 GCGCGCTATGATGTTTCACGACAGTACAAAGTTCCTCTGTATCCACCGTGAAGTCAAT 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
QY 212 CCGGCGTGGTCCATACCCAGAAAGGTCAGCCCATGCTGGAAGATCATCTCAC 271
Db 870 .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACGTTCCGAGCTTCCTTTCCGAGCTGAGCTTCAATTAATCAATCACTACTACGGTTTGC 331
Db 810 .HWCT.KKT.MBHMDS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTGAGCCCTCGGTTCCCTGACGGGAACGAGGATGATCCATTCACACCCAG 386
Db 750 RTYNHM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

RESULT 9
US-10-141-761-394/c
; Sequence 394, Application US/10141761
; Publication No. US2003014832A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141.761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-761-394

Query Match
Best Local Similarity 2.8%; Score 38.6; DB 15; Length 1184;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY 92 CAAGGACCACAAAATGGTGTATAGCACAGCTCCGATGTTTGCACCATGATGCGGG 151
Db 990 C.MT.RMNCYDN.M.....C.D.S...S...B.SHA...B.NY.T.TTBS.YB.SY.. 931
QY 152 GCGCGCTATGATGTTTCACGACAGTACAAAGTTCCTCTGTATCCACCGTGAAGTCAAT 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
QY 212 CCGGCGTGGTCCATACCCAGAAAGGTCAGCCCATGCTGGAAGATCATCTCAC 271
Db 870 .....N.N.YSTNY.Y.DN...NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACGTTCCGAGCTTCCTTTCCGAGCTGAGCTTCAATTAATCAATCACTACTACGGTTTGC 331
Db 810 .HWCT.KKT.MBHMDS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTGAGCCCTCGGTTCCCTGACGGGAACGAGGATGATCCATTCACACCCAG 386
Db 750 RTYNHM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSY.DHBSAB 696

RESULT 8
US-10-140-472-394/c
; Sequence 394, Application US/10140472
; Publication No. US2003013888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140.472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-472-394

Query Match
Best Local Similarity 2.8%; Score 38.6; DB 15; Length 1184;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

QY 92 CAAGGACCACAAAATGGTGTATAGCACAGCTCCGATGTTTGCACCATGATGCGGG 151

```



```

; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C183
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-871-394

Query Match      2.8%; Score 38.6; DB 16; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

Qy 92 CAAGGACCACACAAAATGGTGATAGCAGCTCCGATGTTGGCCACCATGATGGCGG 151
Db 990 C.MT.RMNCYDN.M....C.D..S...S...B..SHEA...B.NY.T.TTBS.YB.SY.. 931
Qy 152 GCGCGGTATGATGTTCCAGCCACAGTACAAAGTTCCTCTGTATCCACCGTGAGGTCA 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
Qy 212 CCGCGGTGGTCCATACCCAGAAAGGGTCAGCCCATGCTCAATCAATCACTACTACGTTGC 331
Db 870 .....N.N.YSTNY.Y.DN...NBAH...B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
Qy 272 ACGTTCCGACTTCCTTCGAGCTGAGCTTCAATCAATCACTACTACTACTACTAGTTGC 331
Db 810 .HWCT.KKT.MBHMDS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
Qy 332 ATTGAGCCCTCGGTTCCCTGACGGGAACGAGGATGATCCATCAACACCCAG 386
Db 750 RTYNHMM.AVG.SK..T....C..C..BT.G.BA.KSKS.TC.MYTSHY.DHSAB 696

RESULT 14
US-10-141-756-394/c
; Sequence 394, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C200
; CURRENT APPLICATION NUMBER: US/10/141,756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-756-394

Query Match      2.8%; Score 38.6; DB 16; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;

Qy 92 CAAGGACCACACAAAATGGTGATAGCAGCTCCGATGTTGGCCACCATGATGGCGG 151
Db 990 C.MT.RMNCYDN.M....C.D..S...S...B..SHEA...B.NY.T.TTBS.YB.SY.. 931
Qy 152 GCGCGGTATGATGTTCCAGCCACAGTACAAAGTTCCTCTGTATCCACCGTGAGGTCA 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
Qy 212 CCGCGGTGGTCCATACCCAGAAAGGGTCAGCCCATGCTCAATCAATCACTACTACTAGTTGC 331
Db 870 .....N.N.YSTNY.Y.DN...NBAH...B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
Qy 272 ACGTTCCGACTTCCTTCGAGCTGAGCTTCAATCAATCACTACTACTACTACTAGTTGC 331
Db 810 .HWCT.KKT.MBHMDS.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
Qy 332 ATTGAGCCCTCGGTTCCCTGACGGGAACGAGGATGATCCATCAACACCCAG 386
Db 750 RTYNHMM.AVG.SK..T....C..C..BT.G.BA.KSKS.TC.MYTSHY.DHSAB 696

RESULT 13
US-10-140-923-394/c
; Sequence 394, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT

```

Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
QY 212 CCGGGGTTGGTTCATACCCAGAAAGGTGAGCCCATGCACTGCAAGAGTCACTCAC 271
Db 870N.N.YSTNY.Y.DN..NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACCTTCGAGACTTCCTTCGAGTCTGAGTCTCAATTAATCAATCACTACTACGGTTTC 331
Db 810 .HWCT.KKT.MBHMD.S.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTCGAGCCCTCGGTTCCCTGACGGGAGGAGGATGATCCATTCAACACCCAG 386
Db 750 RTYNHMM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSHY.DHBSAB 696

RESULT 15

US-10-141-759-394/c
; Sequence 394, Application US/10141759
; Publication No. US20030207361A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C197
; CURRENT APPLICATION NUMBER: US/10/141,759
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 394
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-759-394

Query Match 2.8%; Score 38.6; DB 16; Length 1184;
Best Local Similarity 11.2%; Pred. No. 0.092;
Matches 33; Conservative 82; Mismatches 180; Indels 0; Gaps 0;
QY 92 CAAGGACACACAAATGGTGTATAGCACAGCTCGATGTTGCCACATGATGGCGG 151
Db 990 C.MT.RMNCYDN.M....C.D..S...S...B..SHBA...B.NY.T.TTBS.YB.SY.. 931
QY 152 GCGCGCTATGTTTCAGGCACATACAAAGTTCCTCTGTATCCACCGTGAGGTCATCAT 211
Db 930 .Y.CS.M..SBMCN.DY.YY.ATST..Y..T..SC.HCT.YCA.SCTEM..YSY..CAT 871
QY 212 CCGGGGTTGGTTCATACCCAGAAAGGTGAGCCCATGCACTGCAAGAGTCACTCAC 271
Db 870N.N.YSTNY.Y.DN..NBAH..B.YS.T.TC.NC.NY.AYR.A.A.D..T..G 811
QY 272 ACCTTCGAGACTTCCTTCGAGTCTGAGTCTCAATTAATCAATCACTACTACGGTTTC 331
Db 810 .HWCT.KKT.MBHMD.S.CB.G..CT.CY.B.B.D.HTM..DM..YM...YRAS.T.YGN 751
QY 332 ATTCGAGCCCTCGGTTCCCTGACGGGAGGAGGATGATCCATTCAACACCCAG 386
Db 750 RTYNHMM.AYG.SK..T...C..C..BT.G.BA.KSKS.TC.MYTSHY.DHBSAB 696

OM nucleic - nucleic search, using sw model
Run on: July 28, 2004, 22:11:55 ; Search time 3926 Seconds
(without alignments)
10595.522 Million cell updates/sec

Title: US-10-099-704-1
Perfect score: 1393
Sequence: 1 ccgaagctgagcaatggag.....cgaaaaaaaaaaaaaa 1393

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_fam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	70.8	5.1	557	28	BZ424676
C 2	70.8	5.1	599	28	BZ424729
C 3	70.8	5.1	600	28	BZ424673
C 4	70.8	5.1	602	28	BZ424733

5	46.4	3.3	1201	13	BX381961
6	42.6	3.1	601	28	BZ424674
7	42.6	3.1	601	28	BZ424732
8	41.2	3.0	482	28	BZ424543
9	40	2.9	1019	28	BZ699387
10	39.4	2.8	438	14	CF301206
11	38.4	2.8	602	9	AV609620
12	38.4	2.8	836	10	BE873147
13	38.4	2.8	885	13	BX425603
14	38.4	2.8	1232	12	BM548562
15	38.2	2.7	1190	14	CD508019
16	38.2	2.7	1201	13	BX457923
17	38	2.7	1201	13	BX381961
18	37.8	2.7	397	13	BQ171803
19	37.8	2.7	1201	13	BX382648
20	37.8	2.7	1435	11	BC050967
21	37.4	2.7	361	29	CEG28815
22	37.2	2.7	740	13	BU683350
23	37.2	2.7	854	14	CG528654
24	37	2.7	299	9	AV217652
25	37	2.7	906	12	BG298933
26	36.8	2.6	154	12	BU565098
27	36.8	2.6	284	28	BZ424728
28	36.8	2.6	353	28	BZ424731
29	36.8	2.6	479	12	BJ210477
30	36.8	2.6	489	28	BZ424734
31	36.8	2.6	494	28	BZ424678
32	36.8	2.6	541	28	BZ424665
33	36.8	2.6	588	28	BZ424675
34	36.8	2.6	597	13	BU805984
35	36.8	2.6	1201	9	AL543409
36	36.6	2.6	530	28	AQ634647
37	36.6	2.6	645	13	BQ14025
38	36.6	2.6	742	28	AQ752529
39	36.4	2.6	320	9	AV65857
40	36.4	2.6	1016	14	CF266426
41	36.2	2.6	526	9	AL383443
42	36.2	2.6	637	13	BU895985
43	36	2.6	348	12	BM751760
44	35.8	2.6	588	14	CA814207
45	35.8	2.6	119	12	BU560428

ALIGNMENTS

RESULT 1
BZ424676/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BZ424676 557 bp DNA linear GSS 13-DEC-2002
100017849-5752 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.

BZ424676 GI:26666131
GSS.

Aspergillus terreus
Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

1 (bases 1 to 557)
Askenazi, V., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S.,
Zimmer, D.P., Boers, M.-E., Blomquist, P.R., Martinez, E.J., Sykes, K.,
Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Royer, J.C. and
Madden, K.T.

Integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
Unpublished (2002)
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com

TITLE
JOURNAL
COMMENT

```

FEATURES
  source
    Class: plasmid ends.
    Location/Qualifiers
      1..557
        /organism="Aspergillus terreus"
        /mol_type="genomic DNA"
        /strain="ATCC 20542 (A. terreus Thom, anamorph)"
        /db_xref="taxon:33178"
        /lab_host="Escherichia coli"
        /clone_lib="Aspergillus terreus random genomic DNA clone library"
        /notes="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
  Query Match      5.1%; Score 70.8; DB 28; Length 557;
  Best Local Similarity 49.1%; Pred. No. 0.00022;
  Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

QY 54 CCCTTCTACGCTGCTGAGTCAAGCCCTGAACCTCTCGAACAAGGACACACACAAAATGGTGG 113
   |||
Db 453 CCCTGGAGATATTGGACAAACGACCGGATTCCAAACCAAGACGAGGCTCTGGTGG 394

QY 114 TATAGACACAGCTCGATGTTTGGCCACATGATGGCGGGGCGGGCTATGATGTTACGCA 173
   |||
Db 393 CTAACACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCCTAATGACGTCCTCACTTG 334

QY 174 CAGTACAAAGTCTCTGTATCCACCGTGAGTCAATCCCGCGTGGTGGTCCATACCA 233
   |||
Db 333 CAGTACCAATACCTCAGTCTTCCGCCACCATGTCATTCCTGTTCTAGGGCCCTCTTT 274

QY 234 GAAAAGGTCAGCCCATGCACTGGAAGAGTCACTCACAACGCTTCGAGTCTCTTTTCGAG 293
   |||
Db 273 GCCCAGGAGCACTCCAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 214

QY 294 CTGAGCTTCAATTT-----ACTCCAATCACTACTAGGTTTGATTCAGCGCCCTCGT 347
   |||
Db 213 TTGAGGTCATTTTCAGGAGTCCGGTGAACAGTCCGAATGAGCTGGGGGCCATCGT 154

QY 348 TCCCTGACGGGAACGAAGATGATCCATTCACACCCAGGCAATCAGGCTGTCTCCAG 407
   |||
Db 153 AGCTTTGCTGCTTCGAGAGGATCGTTGAACCACTTCAGGCGGAGAGATTTCTCGAC 94

QY 408 GACTCAAGGCCATGTTTCAGGGCTTGACCTGGAATGTTGATCAATTTCACTAAAGCA 467
   |||
Db 93 AAGCTAGCCATCTTGTACCCACCGTGGACTTGCAGTTGTTCAACACTTCGAGAATGAG 34

QY 468 TTGCTGCTTTTCGAGGAAGAGGCTCGGA 495
   |||
Db 33 TTGCGCATCAACCATGCCGACGGCTGA 6

RESULT 2
BZ424729/c
LOCUS
DEFINITION
  100018692-5746 Aspergillus terreus random genomic DNA clone library
  Aspergillus terreus genomic, genomic survey sequence.
ACCESSION
  BZ424729
VERSION
  BZ424729.1 GI:26666184
SOURCE
  Aspergillus terreus
  Aspergillus terreus
  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
  Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
  1 (bases 1 to 599)
  Askenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S.,
  Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J.,
  Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K.,
  Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and
  Madden,K.T.
  Integrating transcriptional and metabolite profiles to direct the
  engineering of lovastatin-producing strains
  Unpublished (2002)

TITLE
  JOURNAL
  
```

```

COMMENT
  Contact: Zimmer DP
  Microbia, Inc.
  One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
  Tel: 617-621-8322
  Fax: 617-
  Email: dzimmer@microbia.com
  Class: plasmid ends.
  Location/Qualifiers
    1..599
      /organism="Aspergillus terreus"
      /mol_type="genomic DNA"
      /strain="ATCC 20542 (A. terreus Thom, anamorph)"
      /db_xref="taxon:33178"
      /lab_host="Escherichia coli"
      /clone_lib="Aspergillus terreus random genomic DNA clone library"
      /notes="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

FEATURES
  source
    Class: plasmid ends.
    Location/Qualifiers
      1..599
        /organism="Aspergillus terreus"
        /mol_type="genomic DNA"
        /strain="ATCC 20542 (A. terreus Thom, anamorph)"
        /db_xref="taxon:33178"
        /lab_host="Escherichia coli"
        /clone_lib="Aspergillus terreus random genomic DNA clone library"
        /notes="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
  Query Match      5.1%; Score 70.8; DB 28; Length 599;
  Best Local Similarity 49.1%; Pred. No. 0.00022;
  Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

QY 54 CCCTTCTACGCTGCTGAGTCAAGCCCTGAACCTCTCGAACAAGGACACACAAAATGGTGG 113
   |||
Db 453 CCCTGGAGATATTGGACAAACGACCGGATTCCAAACCAAGACGAGGCTCTGGTGG 394

QY 114 TATAGACACAGCTCGATGTTTGGCCACATGATGGCGGGGCGGGCTATGATGTTACGCA 173
   |||
Db 393 CTAACACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCCTAATGACGTCCTCACTTG 334

QY 174 CAGTACAAAGTCTCTGTATCCACCGTGAGTCAATCCCGCGTGGTGGTCCATACCA 233
   |||
Db 333 CAGTACCAATACCTCAGTCTTCCGCCACCATGTCATTCCTGTTCTAGGGCCCTCTTT 274

QY 234 GAAAAGGTCAGCCCATGCACTGGAAGAGTCACTCACAACGCTTCGAGTCTCTTTTCGAG 293
   |||
Db 273 GCCCAGGAGCACTCCAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 214

QY 294 CTGAGCTTCAATTT-----ACTCCAATCACTACTAGGTTTGATTCAGCGCCCTCGT 347
   |||
Db 213 TTGAGGTCATTTTCAGGAGTCCGGTGAACAGTCCGAATGAGCTGGGGGCCATCGT 154

QY 348 TCCCTGACGGGAACGAAGATGATCCATTCACACCCAGGCAATCAGGCTGTCTCCAG 407
   |||
Db 153 AGCTTTGCTGCTTCGAGAGGATCGTTGAACCACTTCAGGCGGAGAGATTTCTCGAC 94

QY 408 GACTCAAGGCCATGTTTCAGGGCTTGACCTGGAATGTTGATCAATTTCACTAAAGCA 467
   |||
Db 93 AAGCTAGCCATCTTGTACCCACCGTGGACTTGCAGTTGTTCAACACTTCGAGAATGAG 34

QY 468 TTGCTGCTTTTCGAGGAAGAGGCTCGGA 495
   |||
Db 33 TTGCGCATCAACCATGCCGACGGCTGA 6

RESULT 3
BZ424673
LOCUS
DEFINITION
  100017849-5764 Aspergillus terreus random genomic DNA clone library
  Aspergillus terreus genomic, genomic survey sequence.
ACCESSION
  BZ424673
VERSION
  BZ424673.1 GI:26666128
SOURCE
  Aspergillus terreus
  Aspergillus terreus
  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
  Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
  1 (bases 1 to 600)
  Askenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S.,
  Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J.,
  Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K.,
  Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and
  Madden,K.T.
  Integrating transcriptional and metabolite profiles to direct the
  engineering of lovastatin-producing strains
  Unpublished (2002)

TITLE
  JOURNAL
  
```

Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
source
1. .600
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

FEATURES
source
1. .600
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 600;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 54 CCCTTCTACGTCGTGAGTCAAGCCCTGAAGCCCTCGAAGCCGACACACAAATGGTGG 113
DB 71 CCTTGAAGATATGGGACAAAGACCGGATTCCTCCAAACCAAGACGAGCTCTGGTG 130
QY 114 TATAGCAGCTCCGATGTTGCCACCATGATGGGGGGCGGCTATGATTTCAAGCA 173
DB 131 CTAACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCAATATGACGTCCACTTG 190
QY 174 CAGTACAGTCTCTGTATCACCCTGAGGTATCATCCCGCGTGGTCCATACCCA 233
DB 191 CAGTACCAATCTCAGTTCTTCCGCCACCATGTCATTCCTGTTCTAGGCCCTCTTT 250
QY 234 GAAAGGGTCAGCCCATGCACTGGAAGATGATCTCACACGCTTCGGACTTCCTTTCCAG 293
DB 251 GCCCAGGACGACTCCAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 310
QY 294 CTGAGCTCAATT-----ACTCCAAATCACTACTAGGTTTGGATTCGAGCCCTCGGT 347
DB 311 TTCAGGCTCAATTTCAGGAGTCCGGTGCAACAGTCCGAATGAGCCTGGGGCCATCGT 370
QY 348 TCCTCAGCGGAACGAAGGATGATCCATTCAACACCGGCAATCAGGCTCTGTTCTCCAG 407
DB 371 AGCTTGTGGCTTGAGCAGAGATCCGTTGAAACAGTTCAGGGCGAGAGAATTTCTGAC 430
QY 408 GACCTCAAGGCCATGTTCCAGGGCTTGACCTGGAATGTTGATCATTTTCACTAAAGCA 467
DB 431 AAGCTAGCCATCTGTATACCAACCGTGGACTTGCACTGTTTCAACACACTTCGAGAATGAG 490
QY 468 TTGGTGTTCGGAGGAAGGCTCGGA 495
DB 491 TTGGCATCAACCATGCCAGCGCTGA 518

RESULT 4
BZ424733 602 bp DNA linear GSS 13-DEC-2002
LOCUS BZ424733
DEFINITION 100018692-5758 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BZ424733
VERSION BZ424733.1 GI:26666188
KEYWORDS GSS.
SOURCE Aspergillus terreus

ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 602)
AUTHORS Akenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S., Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J., Sykes,K., Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
source
1. .602
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

FEATURES
source
1. .602
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 602;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 54 CCCTTCTACGTCGTGAGTCAAGCCCTGAAGCCCTCGAAGCCGACACACAAATGGTGG 113
DB 72 CCTTGAAGATATGGGACAAAGACCGGATTCCTCCAAACCAAGACGAGCTCTGGTG 131
QY 114 TATAGCAGCTCCGATGTTGCCACCATGATGGGGGGCGGCTATGATTTCAAGCA 173
DB 132 CTAACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCCAATATGACGTCCACTTG 191
QY 174 CAGTACAGTCTCTGTATCACCCTGAGGTATCATCCCGGCTGGGTCCATACCCA 233
DB 192 CAGTACCAATCACTACTAGGTTTCCGCCACCATGTCATTCCTGTTCTAGGCCCTCTTT 251
QY 234 GAAAGGGTCAGCCCATGCACTGGAAGAGTCACTCACACGCTTCGAGCTTCCTTTCCAG 293
DB 252 GCCCAGGACGACTCCAAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 311
QY 294 CTGAGCTCAATT-----ACTCCAAATCACTACTAGGTTTGGATTCGAGCCCTCGGT 347
DB 312 TTCAGGCTCAATTTCAGGAGTCCGGTGCAACAGTCCGAATGAGCCTGGGGCCATCGT 371
QY 348 TCCTCAGCGGAACGAAGGATGATCCATTCAACACCGGCAATCAGGCTCTGTTCTCCAG 407
DB 372 AGCTTGTGGCTTGAGCAGAGATCCGTTGAAACAGTTCAGGGCGAGAGAATTTCTGAC 431
QY 408 GACCTCAAGGCCATGTTCCAGGGCTTGACCTGGAATGTTGATCATTTTCACTAAAGCA 467
DB 432 AAGCTAGCCATCTGTATACCAACCGTGGACTTGCACTGTTTCAACACACTTCGAGAATGAG 491
QY 468 TTGGTGTTCGGAGGAAGGCTCGGA 495
DB 492 TTGGCATCAACCATGCCAGCGCTGA 519

RESULT 5
BX381961
LOCUS BX381961
DEFINITION 100018692-5758 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:26666188
KEYWORDS GSS.
SOURCE Aspergillus terreus

ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 602)
AUTHORS Akenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S., Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J., Sykes,K., Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
source
1. .602
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

FEATURES
source
1. .602
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 602;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 54 CCCTTCTACGTCGTGAGTCAAGCCCTGAAGCCCTCGAAGCCGACACACAAATGGTGG 113
DB 72 CCTTGAAGATATGGGACAAAGACCGGATTCCTCCAAACCAAGACGAGCTCTGGTG 131
QY 114 TATAGCAGCTCCGATGTTGCCACCATGATGGGGGGCGGCTATGATTTCAAGCA 173
DB 132 CTAACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCCAATATGACGTCCACTTG 191
QY 174 CAGTACAGTCTCTGTATCACCCTGAGGTATCATCCCGGCTGGGTCCATACCCA 233
DB 192 CAGTACCAATCACTACTAGGTTTCCGCCACCATGTCATTCCTGTTCTAGGCCCTCTTT 251
QY 234 GAAAGGGTCAGCCCATGCACTGGAAGAGTCACTCACACGCTTCGAGCTTCCTTTCCAG 293
DB 252 GCCCAGGACGACTCCAAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 311
QY 294 CTGAGCTCAATT-----ACTCCAAATCACTACTAGGTTTGGATTCGAGCCCTCGGT 347
DB 312 TTCAGGCTCAATTTCAGGAGTCCGGTGCAACAGTCCGAATGAGCCTGGGGCCATCGT 371
QY 348 TCCTCAGCGGAACGAAGGATGATCCATTCAACACCGGCAATCAGGCTCTGTTCTCCAG 407
DB 372 AGCTTGTGGCTTGAGCAGAGATCCGTTGAAACAGTTCAGGGCGAGAGAATTTCTGAC 431
QY 408 GACCTCAAGGCCATGTTCCAGGGCTTGACCTGGAATGTTGATCATTTTCACTAAAGCA 467
DB 432 AAGCTAGCCATCTGTATACCAACCGTGGACTTGCACTGTTTCAACACACTTCGAGAATGAG 491
QY 468 TTGGTGTTCGGAGGAAGGCTCGGA 495
DB 492 TTGGCATCAACCATGCCAGCGCTGA 519

RESULT 5
BX381961
LOCUS BX381961
DEFINITION 100018692-5758 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:26666188
KEYWORDS GSS.
SOURCE Aspergillus terreus

ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 602)
AUTHORS Akenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S., Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J., Sykes,K., Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
source
1. .602
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

FEATURES
source
1. .602
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 5.1%; Score 70.8; DB 28; Length 602;
Best Local Similarity 49.1%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 222; Indels 6; Gaps 1;
QY 54 CCCTTCTACGTCGTGAGTCAAGCCCTGAAGCCCTCGAAGCCGACACACAAATGGTGG 113
DB 72 CCTTGAAGATATGGGACAAAGACCGGATTCCTCCAAACCAAGACGAGCTCTGGTG 131
QY 114 TATAGCAGCTCCGATGTTGCCACCATGATGGGGGGCGGCTATGATTTCAAGCA 173
DB 132 CTAACACCGCCCTCTGCTCAACGAATTTCTGGCGAGTGCCAATATGACGTCCACTTG 191
QY 174 CAGTACAGTCTCTGTATCACCCTGAGGTATCATCCCGGCTGGGTCCATACCCA 233
DB 192 CAGTACCAATCACTACTAGGTTTCCGCCACCATGTCATTCCTGTTCTAGGCCCTCTTT 251
QY 234 GAAAGGGTCAGCCCATGCACTGGAAGAGTCACTCACACGCTTCGAGCTTCCTTTCCAG 293
DB 252 GCCCAGGACGACTCCAAACTTCGCCAGCAGACTCAGCAAGCAGCGCCACCTCTTGAT 311
QY 294 CTGAGCTCAATT-----ACTCCAAATCACTACTAGGTTTGGATTCGAGCCCTCGGT 347
DB 312 TTCAGGCTCAATTTCAGGAGTCCGGTGCAACAGTCCGAATGAGCCTGGGGCCATCGT 371
QY 348 TCCTCAGCGGAACGAAGGATGATCCATTCAACACCGGCAATCAGGCTCTGTTCTCCAG 407
DB 372 AGCTTGTGGCTTGAGCAGAGATCCGTTGAAACAGTTCAGGGCGAGAGAATTTCTGAC 431
QY 408 GACCTCAAGGCCATGTTCCAGGGCTTGACCTGGAATGTTGATCATTTTCACTAAAGCA 467
DB 432 AAGCTAGCCATCTGTATACCAACCGTGGACTTGCACTGTTTCAACACACTTCGAGAATGAG 491
QY 468 TTGGTGTTCGGAGGAAGGCTCGGA 495
DB 492 TTGGCATCAACCATGCCAGCGCTGA 519

RESULT 5
BX381961
LOCUS BX381961
DEFINITION 100018692-5758 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:26666188
KEYWORDS GSS.
SOURCE Aspergillus terreus

```
DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cdna
clone CSODI072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:304533007
EST. EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI072CC03NP1.
FEATURES             Location/Qualifiers
     source            1..1201
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="CSODI072YF05"
                        /tissue_type="PLACENTA COT 25-NORMALIZED"
                        /note="1st strand cDNA was primed with a NotI-oligo(dT)
                        primer. Five prime end enriched, double-strand cDNA was
                        digested with Not I and cloned into the Not I and EcoR V
                        sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match          3.3%; Score 46.4; DB 13; Length 1201;
Best Local Similarity 4.1%; Pred. No. 22;
Matches 29; Conservative 246; Mismatches 433; Indels 0; Gaps 0;

QY 88 CGACAGAGGACACAAAATGGTGTATAGACAGCTCCGATGTTGCCACCATGATGG 147
DB 381 CNVACANNKNNKKKACNANNNNKCKMMNNNNKCKKCKMNNKNNKNNNNKNNNNK 440
QY 148 CGGGGCGCGGTATGATGTTACGACAGTCAAGTTCCTCTGTATCCACCGTGAGGTCA 207
DB 441 KNNNNNCNANNNNNNKNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 500
QY 208 TCATCCGGGGTGGGTCCATACCAGAAAAGGTCAGCCCATGCACTGGAAGATCATC 267
DB 501 KNNNNNNNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 560
QY 268 TCACAGCTTCGGACTTCCTTCGAGCTCAGCTTCAATTACTCCAAATCACTACTACGT 327
DB 561 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 620
QY 328 TTGCATTCGAGCCCTCGGTTCCTTCGAGCGGAAACGAAGATGATCAATCAACACCCAGG 387
DB 621 KKKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 680
QY 388 CAATCAGGCTGTCTCCAGGACCTCAAGCCATGTTCCAGGGCTTGACCTGGATGGT 447
DB 681 KKKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 740
QY 448 TCGATCAATTTCAATAAGCATTTGGTTCGTTTCGAGGAGAGAGGCTCGGACTTCGTAGATC 507
DB 741 MNNNCNNKNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 800
QY 508 GAGATATTGAGATCCCGGTCTTCAAGACACAGAAACAACTGCGACCGCATCTGAGCCAT 567
DB 801 CNVCKKKKKKKKKKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 860
QY 568 CTGGCGATATTGTCTTGAAGACCTACATCTACCCGCGGATCAAGTCGATCGCGACCGGGA 627
```

```
DB 861 MNNNNKAMNKEDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 920
QY 628 CCCCAAAAGAGAGACTCATGTTTGAACCAATCAAGGCTGCGGACCAAGTTTGGCAAAAGTTG 687
DB 921 MNCCKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 980
QY 688 CCATCCATCGGCAATCTCTGAGGAGTTTATAGCTGAGCGAGCACCACCTCTCTCGGCC 747
DB 981 KKVCMGCDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1040
QY 748 ACTTCTCTCATGCGATTTGGTCAAGCGCTCCGAGTCCCGAATCAAGG 795
DB 1041 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1088

RESULT 6
BZ424674
LOCUS BZ424674
DEFINITION 100017849-5762 Aspergillus terreus random genomic DNA clone library
Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BZ424674
VERSION BZ424674.1 GI:26666129
KEYWORDS GSS.
SOURCE Aspergillus terreus
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Askenazi, M., Boegers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S.,
Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J.,
Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K.,
Tobin, J., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and
Madden, K.T.
TITLE Integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
JOURNAL Unpublished (2002)
COMMENT Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
FEATURES             Location/Qualifiers
     source            1..601
                        /organism="Aspergillus terreus"
                        /mol_type="genomic DNA"
                        /strain="ATCC 20542 (A. terreus Thom, anamorph)"
                        /db_xref="taxon:33178"
                        /lab_host="Escherichia coli"
                        /clone_lib="Aspergillus terreus random genomic DNA clone
                        library"
                        /note="Vector: pZER-OTM-2; Site 1: Sau3A; Site 2: BamHI;
                        Sau3A genomic fragments ligated into BamHI digested
                        pZER-OTM-2"
ORIGIN
Query Match          3.1%; Score 42.6; DB 28; Length 601;
Best Local Similarity 52.5%; Pred. No. 1.5e+02;
Matches 93; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 54 CCCTTCTACGTGCTGAGTCAAGCCCTGACCTCTCGAACAGGACACACAAATGCTGG 113
DB 384 CCTTGAAGATATTGGGACAAACGACCGGATTCCTCCAAACCAAGACGAGGAGCTCTGCTGG 443
QY 114 TATAGCACACTCCGATGTTTGCACCATGATCGGGGGCGGCTATGATGTTCAAGCA 173
DB 444 CTAAACACGCCCTCTGCTCAACGATTTCTGCGCGAGTGCAATATGCTCCACTTG 503
QY 174 CAGTACAGTTCTCTGTATCCACCGTGAGGTCAATCATCCCGCGGTGGTGGTCCATAC 230
DB 504 CAGTACCAATACCTACGTTCTTCCGCCACCATGTCATCTCTGTTCTAGGCGCCCTTC 560
```

AUTHORS Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and Madden, K.T.

TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains

JOURNAL COMMENT Unpublished (2002)
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
1. .482
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

FEATURES source
1. .482
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 3.0%; Score 41.2; DB 28; Length 482;
Best Local Similarity 49.5%; Pred. No. 3.1e+02;
Matches 106; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 282 CTTCCTTCGAGCTGAGCTTCAATTACTCCAAATCACTACTACCGTTTGCATTCGAGCCC 341
Db 468 CTTGATTTACCGTCATTTCCAGGAGTCCGGTCCACAGCTCCGAATGAGCCTGGGGCC 409
QY 342 CTCGGTTCCCTGACGGAAAGAGATGATCATTCAACACCCAGGCAATCAGGCCTGTT 401
Db 408 ATCGTAGCTTGTGCTTCGAGGATCCGTTGAACCAAGTTCAGGGCGAGAGAGTT 349
QY 402 CTCGAGGACCTCAAGGCGATGTTCCAGGGCTTCAGCTGGAATGTTGATTCATTTCACT 461
Db 348 CTCGACAGCTAGCCATCTTGTACCCAACTGAGCTTGCTGTTTCAACACTTCGAG 289
QY 462 AAAGCATTTGCTGTTTCGGAGGAAGAGGCTCGGA 495
Db 288 AATGAGTTGGCATCAACCAATGCCGACGCGCTGA 255

RESULT 9
BZ699387/c
LOCUS PUCEG32TD ZM.0.6.1.0 KB Zea mays genomic clone ZMMBTa126F15,
DEFINITION genomic survey sequence.
ACCESSION BZ699387
VERSION BZ699387.1 GI:28419234
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1. (bases 1 to 1019)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843

RESULT 7
BZ424732
LOCUS 601 bp DNA linear GSS 13-DEC-2002
DEFINITION Aspergillus terreus random genomic DNA clone library
ACCESSION BZ424732
VERSION BZ424732.1 GI:26666187
KEYWORDS GSS.
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Askenazi, M., Driggers, E.M., Holtzman, D.A., Norman, T.C., Iverson, S., Zimmer, D.P., Boers, M.E., Blomquist, P.R., Martinez, E.J., Monreal, A.W., Feibelman, T.P., Mayorga, M.E., Maxon, M.E., Sykes, K., Tobin, J., Cordero, E., Salama, S.R., Trueheart, J., Royer, J.C. and Madden, K.T.

TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains

JOURNAL COMMENT Unpublished (2002)
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
1. .601
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:33178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"

ORIGIN
Query Match 3.1%; Score 42.6; DB 28; Length 601;
Best Local Similarity 52.5%; Pred. NO. 1.5e+02;
Matches 93; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 54 CCCTTCTACGTGCTGAGTCAAGCCCTGAACCTTCGAACAGGACCCACACAAATGGTGG 113
Db 383 CCTGGAAGATATTGGCAACAGCAGCGGATTCGCGAGTGCCAAATACGCTCCACTTG 442
QY 114 TATAGCACAGCTCCGATGTTGCCACCATGATGCGGGGCGCGCTATGATGTTCAAGCA 173
Db 443 CTAACACCGCCCTCTGCTCAAGCAATTTCTGCGAGTGCCAAATACGCTCCACTTG 502
QY 174 CAGTACAGTCTCTGATATCCACCGTGAGGTATCATCCCGCGCTGGTCCATAC 230
Db 503 CAGTACCAATACCTCAGTTCCTTCGCGCCACCATGTCTATCTCTTAGGGCCCTTC 559

RESULT 8
BZ424543/c
LOCUS 482 bp DNA linear GSS 13-DEC-2002
DEFINITION Aspergillus terreus random genomic DNA clone library
ACCESSION BZ424543
VERSION BZ424543.1 GI:26665998
KEYWORDS GSS.
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 482)

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

FEATURES

source

1. 1019
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBR126F15"
/clone_lib="ZM 0.6-1.0 kb"
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN

Query Match 2.9%; Score 40; DB 28; Length 1019;
Best Local Similarity 58.3%; Pred. No. 4.7e+02;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 1029 GCACCTCCCGAACACAGATCTATTCCTGCTTTGGGAGAACACCAACCATCGCG 1088
Db 179 GCACCCACCGAACACAAATCTATTATCCGATTTTCAAAACACGACACAATTTGTGAT 120
Qy 1089 GAAGGATGGCCACCTTCCTTGGAGACGAGGTTGGGGTGGCTTAAGAGCTATCCA 1148
Db 119 GAACAAATGCCCAATCTTTAGTAGTGATGGTGTGGTCTTTCCGGGTGCCAATCA 60

RESULT 10

CF301206/c

LOCUS 438 bp mRNA linear EST 15-AUG-2003
DEFINITION 7LEAF--05-P13.b1 Rice leaf plasmid cDNA library II (7LEAF) Oryza
sativa cDNA clone 7LEAF--05-P13, mRNA sequence.
ACCESSION CF301206
VERSION
KEYWORDS EST.
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 438)
AUTHORS Kim J.S., Jun K.M., Cheong P.J., Kim M.J., Lee T.H., Shin Y.C.,
Song S.I., Kim J.K., Kim Y.-K. and Nahm B.H.
TITLE Large-scale Sequencing Analysis of Rice ESTs
JOURNAL Unpublished (2003)
COMMENT Contact: Nahm B.H.
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of Bioscience and Bioinformatics, Myongji University
Yongin, Gyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@bio.com, bhnahm@bio.myongji.ac.kr.

FEATURES

source

1. 438
/organism="Oryza sativa"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:4530"
/clone="7LEAF--05-P13"
/tissue_type="leaf"
/dev_stage="7 days after germination"
/clone_lib="Rice leaf plasmid cDNA library II (7LEAF)"
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; mRNA was capped
with oligoribonucleotides and then used as templates for
RT-PCR."

ORIGIN

Query Match 2.8%; Score 39.4; DB 14; Length 438;
Best Local Similarity 67.9%; Pred. No. 7.3e+02;
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1313 TGGCCACAATCCTTAGGACTAGTTTATCCCTTATCTATCCATCCGTTGAATGTTGG 1372
Db 156 TGGCCCAACCCCTTGGGAAGGGTGGTCTTTTTTTTAAATTTGAATTTGTTGG 97
Qy 1373 TCGAAAAAATAAAAAAAAAA 1393
Db 96 GGGAAAAAATAAAAAAAAAA 76

RESULT 11

AV609620

LOCUS 602 bp mRNA linear EST 28-NOV-2001
DEFINITION AV609620 Bos taurus lung fetus Bos taurus cDNA clone EILU025E08 5',
mRNA sequence.
ACCESSION AV609620 GI:9745290
VERSION
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 602)
AUTHORS Takasuga A., Hirotsune S., Itoh R., Jitohsano A., Suzuki H., Aso H.
and Sugimoto Y.
TITLE Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
PUBMED 11713328
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusuga@ccoc.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES

source

1. 602
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="EILU025E08"
/tissue_type="lung"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus lung fetus"
/notes="Vector: pZLI; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"

ORIGIN

Query Match 2.8%; Score 38.4; DB 9; Length 602;
Best Local Similarity 47.6%; Pred. No. 1.1e+03;
Matches 111; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
Qy 10 GAGCAATGGAGATCTCCAGAAAGCAGACACTGTGCGAAAGCCCTTCTAGTGCTGA 69
Db 70 GACCATNGGTGGGCTCCCGCGGCCCACTGCTCTCTAGGCATCCCTAGATCGTCG 129
Qy 70 GTCAAGCCCTGAACTCTCTGAAACAAGGACACACAAAATGGTGGTATAGCACAGCTCCGA 129
Db 130 CCTGCGCCCTGGCCCTGGCGGTGAGCCCGCGCAGCGGCGCTAGGAGGTCGCC 189
Qy 130 TGTTCGCCACCATGATGGCGGGGCGCGGTATGATTTTCCGACAGTACAGTTCCTCT 189
Db 190 TGTGCGCGGCTGATGGAGGCGGACGTCATGAGGAGGCGGTGACAGAGGCGCTGCT 249
Qy 190 GTATCCACCGTAGGTCATCATCCCGCGGTGGGTGCTCATACCCAGAAAAGGTT 242

Db 250 TTGGCGTACGAGTTCAACAAGCGGAGCACTGAGCTTACCAGCGCGT 302

RESULT 12
BE873147
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BE873147 836 bp mRNA linear EST 20-OCT-2000
601451658F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3855327 5',
mRNA sequence.
BE873147
BE873147.1 GI:10321923
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 836)
NIH-MGC <http://imgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM9582 row: d column: 16
High quality sequence stop: 679.
Location/Qualifiers
1..836
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3855327"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 65"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

FEATURES
source

Query Match 2.8%; Score 38.4; DB 10; Length 836;
Best Local Similarity 48.6%; Pred. No. 1e+03;
Matches 105; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 652 ACGCAATCAAGGTCGCGCAAGTTTGGCAAGTTCGCACTCCACTCGCAATCTCGAGG 711
|||||
Db 436 ACGCCATGAATGGCGCGAGATCACAGACAAAGCTGGGCTGCACCTACGACCAAGG 495
|||||
QY 712 AGTTTATAGCTGAGCGAGACCCACCTCTCGGCGCACTTTCTCTCATGCGATTTGGTCA 771
|||||
Db 496 AACTGTGTACATTAGGCCACCTGGCCACAGCGCGGAGCGGCTCTATGAGNACTGACT 555
|||||
QY 772 AGCGGTCGAGTCCCGAATCAAGGTCTACTGTATGGAACGCCAGCTCGACCTGGCTCCA 831
|||||
Db 556 GGCTGTCCATCAGCTCCGGAACAGAAAGTGAACGGACGCCGCCCTCTCTCCTCCTCT 615
|||||
QY 832 TCGNAGTATTGGACTCTCAACGGCGCGGAAAG 867
|||||
Db 616 TGCCTCTGTCTTACTCTCATGTGGCAACCGTGCG 651
|||||

RESULT 13
BX425603/c
LOCUS
DEFINITION
ACCESSION
VERSION

BX425603 885 bp mRNA linear EST 15-MAY-2003
BX425603 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
ClonB0222A07 3-PRIME, mRNA sequence.
BX425603
BX425603.1 GI:30770486

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 885)
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CL08B022ZA07Pp1.
Location/Qualifiers
1..885
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ClonB022ZA07"
/tissue_type="NEUROBLASTOMA"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

FEATURES
source

Query Match 2.8%; Score 38.4; DB 13; Length 885;
Best Local Similarity 2.0%; Pred. No. 1e+03;
Matches 9; Conservative 142; Mismatches 294; Indels 0; Gaps 0;

QY 349 CCCTGACGGGAACGAAGGATGATCCATCAACACCCAGGCAATCAGGCTGTCTCCAGG 408
|||||
Db 803 MNN 744
|||||
QY 409 ACCTCAAGGCCATGGTTCAGGGCTTGACCTGGGAATGGTTTCGATCATTTCTCAAAAGCAT 468
|||||
Db 743 MNN 684
|||||
QY 469 TGGTCTGTTTCGGAGAGAGGCTCGGACTCTGTAGTCGAGATATTGAGATCCCGTCT 528
|||||
Db 683 KGNKKNKKKKGGKGNKNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 624
|||||
QY 529 TCAAGACACAGACAAACAACTGGCAGCGGATCTGGAGCCATCTGGCGATATTGCTTGAAGA 588
|||||
Db 623 ATNN 564
|||||
QY 589 CCTACATCTACCCGCGGATCAAGTCGATCGGACCGGACCCCAAAAGAGAGACTCATGT 648
|||||
Db 563 NNN 504
|||||
QY 649 TTGACGCAATCAAGGCTGCCGACAAGTTTGCAAGTTGGCCACTCCACTGCAATCTCG 708
|||||
Db 503 NNN 444
|||||
QY 709 AGGAGTTTATAGCTGAGGAGCACCCACCTCTCGGCGCACTTTCTCTCATGCGATTGG 768
|||||
Db 443 NNN 384
|||||
QY 769 TCAAGCCGCTCGAGTCCCGCAATCAA 793
|||||
Db 383 NNN 359
|||||

RESULT 14
BM548562
LOCUS
DEFINITION

BM548562 1232 bp mRNA linear EST 20-FEB-2002
AGENCOURT_6573473 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5732079

```

5', mRNA sequence.
ACCESSION BM548562
VERSION BM548562.1 GI:18783222
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 1232)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12733 row: j column: 16
High quality sequence stop: 365.
FEATURES             source
Location/Qualifiers
1..1232
/mol_type="mRNA"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5732079"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"
/notes="Organ: Brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destoyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."

ORIGIN
Query Match 2.8%; Score 38.4; DB 12; Length 1232;
Best Local Similarity 47.2%; Pred. No. 9.7e+02;
Matches 117; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 748 ACATTCTCTCATCGATTGGTCAACGCGTCCGAGTCCGAGATCAAGGTCTACTGTATGG 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 ACCTGGCCACTCGAGGTGGAGCTGTGGCGGAGGCGCTTCCGTGGCTTGGGCGCGCC 286
QY 808 AACGCCAGCTCGACCTGGGCTCCATCGAAGGTATTGGACTCTCAACGGGCGACGGAACG 867
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 TGTCTCTGTCTCACTTGGCTCTCAACCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTG 346
QY 868 ATCCAGAGACATGGATGCTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 927
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 GGCTGGGCTGATGCGCGCGGCTGAGCTGGAGGGGAAACGACTGGACGAGCTGCCACCGC 406
QY 928 CGGAGAGTCTGTCTCCACTGCGCACTGCTTTTACGCGCGGTACCTCACCGAGGAGC 987
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
407 CGACAGATCTCGAATCGGGCGCCATACCCCAACCCACTTGCCTCATCGGCGCCACCC 466
QY 988 AGTCTCCC 995
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
467 AACTACCC 474

RESULT 15
CD508019/c
LOCUS
DEFINITION CD488-E02.y1d-s SHGC-CDA Gasterosteus aculeatus cDNA clone
CD488-E02 3', mRNA sequence.
ACCESSION CD508019

5', mRNA sequence.
CD508019.1 GI:31438588
EST.
Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
REFERENCE
1 (bases 1 to 1190)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthopterygii; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
Kingsley,D.M., Reichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 88
High quality sequence start: 15
High quality sequence stop: 876.
FEATURES             source
Location/Qualifiers
1..1190
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Galinas river, CA"
/db_xref="taxon:69293"
/clone="CDA88-E02"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(adapter); Site 2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-CDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

ORIGIN
Query Match 2.7%; Score 38.2; DB 14; Length 1190;
Best Local Similarity 53.7%; Pred. No. 1.1e+03;
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1103 CTTCTTTGAGCAGAGGTTGGGCTGGCTTGGCTAAAGAGCTATCCAGCGGATTGGCATC 1162
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1005 CTTGCTTTACAGCAGACATGCGGAGTACCGGAGCGCGTTGTCGCCAGAGTTGGACGC 946
QY 1163 CTACTATCCCGATGTGACCTGCAGACCGCAATACCTGCAGCGGTGATCTCTCTC 1222
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
945 CTTGGAAGCGGAGCGGCGGCTGACCGCGCTGCTGCTGAGCGCTGAGGCGCTGGACCGAT 886
QY 1223 TTCAAGGGGAAAAAACCCTACATGAG 1249
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
885 CAAAGAGAGGTGGAGCGCTACCTGAG 859

Search completed: July 29, 2004, 01:17:22
Job time : 3933 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:33:31 ; Search time 54 Seconds
(without alignments)
2286.542 Million cell updates/sec

Title: US-10-099-704-2
Perfect score: 2322
Sequence: 1 MEISKAAATLLPKPFVLSQ.....PEAFSAAQAEVAMCHDGNP 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2322	100.0	437	3 AAY96961	Aay96961 A. oryzae
2	907.5	39.1	448	5 AAU76412	Aau76412 Dimethylyla
3	904.5	39.0	450	5 AAU76413	Aau76413 Dimethylyla
4	103.5	4.5	880	6 AAO27492	Aao27492 Balsam pe
5	102	4.4	521	5 AAM50327	Aam50327 Human nuc
6	102	4.4	606	5 AAO17868	Aao17868 Pyrin dom
7	102	4.4	635	6 ABO07113	Abc07113 Novel hum
8	101	4.3	878	4 AAB11500	Aab11500 C. sativu
9	101	4.3	878	4 AAB86036	Aab86036 Cucumbe
10	99.5	4.3	503	7 ADE54304	Ades4304 Rat Prote
11	99.5	4.3	503	7 ADE54292	Ades4292 Rat Prote
12	99.5	4.3	503	7 ADE54296	Ades4296 Rat Prote
13	99.5	4.3	503	7 ADE60548	Ades60548 Rat Prote
14	99.5	4.3	503	7 ADE62266	Ades62266 Rat Prote
15	99.5	4.3	503	7 ADE54300	Ades4300 Rat Prote
16	99.5	4.3	1210	7 ADE08065	Ades08065 Novel pro
17	99.5	4.3	3234	5 ABP70132	Abp70132 Human NOV
18	98.5	4.2	629	4 ABG04518	Abg04518 Novel hum
19	98.5	4.2	769	6 ABU38122	Abu38122 Protein e
20	97	4.2	1939	5 ABG97437	Abg97437 A. orient
21	96.5	4.2	2627	6 ABR39483	Abp39483 L. cuprin
22	96.5	4.2	3333	6 ABR39482	Abp39482 L. cuprin
23	95.5	4.1	1171	4 AAU32421	Aau32421 Novel hum
24	95	4.1	249	6 ABP78985	Abp78985 N. gonorr
25	95	4.1	604	3 AAY57087	Aay57087 Mouse ret

26	94.5	4.1	317	2	AAW17105	Aaw17105 Amino aci
27	94.5	4.1	317	2	AAW77838	Aaw77838 Multi-fun
28	94.5	4.1	320	2	AAW17115	Aaw17115 Amino aci
29	94.5	4.1	320	2	AAW77848	Aaw77848 Multi-fun
30	94.5	4.1	329	2	AAW17132	Aaw17132 Amino aci
31	94.5	4.1	329	2	AAW77898	Aaw77898 Multi-fun
32	94.5	4.1	430	2	AAY35283	Aay35283 Chlamydia
33	94.5	4.1	430	6	ABU26782	Abu26782 Protein e
34	94.5	4.1	453	5	ABP47775	Abp47775 Protein #
35	94.5	4.1	879	4	ABG25755	Abg25755 Novel hum
36	94.5	4.1	879	4	ABG28383	Abg28383 Novel hum
37	94	4.0	3913	6	ABM67350	Abm67350 Photorhab
38	93	4.0	302	2	AAW17104	Aaw17104 Amino aci
39	93	4.0	302	2	AAW77837	Aaw77837 Multi-fun
40	93	4.0	305	2	AAW17114	Aaw17114 Amino aci
41	93	4.0	305	2	AAW77847	Aaw77847 Multi-fun
42	93	4.0	833	6	ABU00599	Abu00599 S. pneumo
43	93	4.0	2492	5	AAE18302	Aae18302 Venezuela
44	92.5	4.0	585	6	ABU46996	Abu46996 Protein e
45	92.5	4.0	686	4	ABG25945	Abg25945 Novel hum

ALIGNMENTS

RESULT 1
AAV96961
ID AAY96961 standard; protein; 437 AA.
XX
AC AAY96961;
XX
DT 31-OCT-2000 (first entry)
XX
DE A. oryzae dimethylallyl-cycloacetyl-L-tryptophan synthase.
XX
KW Dimethylallyl-cycloacetoacetyl-L-tryptophan synthase; DCAT-S; beta-CPA;
KW toxin production; elimination; heterologous protein production;
KW filamentous fungi.
XX
OS Aspergillus oryzae.
XX
FN WO200039322-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-DK000726.
XX
PR 23-DEC-1998; 98DK-00001726.
PR 27-MAY-1999; 99DK-00000745.
(NOVO) NOVO NORDISK AS.
Christensen BE, Mollgaard H, Kaasgaard S, Lehmebeck J;
WPI; 2000-452411/39.
N-PSDB; AAA51712.

Producing a polypeptide of interest such as a hormone or enzyme, comprising cultivating a mutant of a parent Aspergillus cell which produces less of at least one toxin of interest compared to the parent cell under the same conditions.

Claim 34; Page 61-62; 66pp; English.

The A. oryzae dimethylallyl-cycloacetoacetyl-L-tryptophan synthase (DCAT-S) is involved in the synthesis of beta-CPA, from cyclo-acetoacetyl-L-tryptophan and dimethylallylpyrophosphate, by its homology to a dimethylallyltryptophan synthase (DMAT-S) from Claviceps purpurea. Aspergillus host cells having a modification in the DCAT-S gene, leading to reduced or eliminated toxin production, are useful for expression of heterologous polypeptides of interest. Other toxins which may be reduced or eliminated comprise kojic acid, 3-nitropropionic acid, emodin, etc. The DCAT-S gene can be used to identify and disrupt similar genes in

CC other filamentous fungal host strains such as Trichoderma, Penicillium
 CC and Fusarium
 XX
 SQ Sequence 437 AA;
 Query Match 100.0%; Score 2322; DB 3; Length 437;
 Best Local Similarity 100.0%; Pred. No. 1.3e-214;
 Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEISKKAATLLPPFYVLSQALNSKDHDKWYSTAPFATWAGAGYDVHQAQYKFLCI 60
 Db 1 MEISKKAATLLPPFYVLSQALNSKDHDKWYSTAPFATWAGAGYDVHQAQYKFLCI 60
 QY 61 HREVIIPALGPYKPGQPMHWKSHLTFGLPFELSNYSKSLRFAFEPLGSLTGTKDDP 120
 Db 61 HREVIIPALGPYKPGQPMHWKSHLTFGLPFELSNYSKSLRFAFEPLGSLTGTKDDP 120
 QY 121 FNTQAIRPVLDLQKAMVPGDLDFWDFHTKALVVSBEAEARTLLDRDIEIPVKTQNKLA 180
 Db 121 FNTQAIRPVLDLQKAMVPGDLDFWDFHTKALVVSBEAEARTLLDRDIEIPVKTQNKLA 180
 QY 181 DLSPSGDIVLKYIYPRIKSIATGTPKERMFDIAKAADKFGKVAATPLALEEPIAERAP 240
 Db 181 DLSPSGDIVLKYIYPRIKSIATGTPKERMFDIAKAADKFGKVAATPLALEEPIAERAP 240
 QY 241 TLGHFLSCDLVKPSESRIKVCWQERQDLASIEGIWTLNGRNDPDTLGLDALRELWQ 300
 Db 241 TLGHFLSCDLVKPSESRIKVCWQERQDLASIEGIWTLNGRNDPDTLGLDALRELWQ 300
 QY 301 LLPVTEGLCPNCFEYEPGTSPOQLPFIINFTLSPKALPEQIYEPAPGQNDKTIAG 360
 Db 301 LLPVTEGLCPNCFEYEPGTSPOQLPFIINFTLSPKALPEQIYEPAPGQNDKTIAG 360
 QY 361 LATFFESRGWGLAKSPADLASYPDVLQTAHQLQAWISFSYKGGKPYMSVYLHTFEA 420
 Db 361 LATFFESRGWGLAKSPADLASYPDVLQTAHQLQAWISFSYKGGKPYMSVYLHTFEA 420
 QY 421 FSAAAEVAMCHDGNP 437
 Db 421 FSAAAEVAMCHDGNP 437

RESULT 2
 AAU76412
 ID AAU76412 standard; protein; 448 AA.
 XX
 AC AAU76412;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Dimethylallyltryptophan synthase (DmaW) version #1.
 XX
 KW Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 KW endophyte; forage; pasture; turf; land reclamation; soil conservation;
 KW ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 KW Balansia; Balansioopsis; Echinodopsis; Atkinsonella; Myriogenospora;
 KW Neotyphodium; Parepichloe; knockout; antisense technology.
 XX
 OS Neotyphodium coenophialum.
 XX
 FN US6335188-B1.
 XX
 PD 01-JAN-2002.
 XX
 XX 03-MAR-2000; 2000US-00518657.
 XX
 PR 22-MAR-1999; 99US-0125490P.
 XX
 XX (KENT) UNIV KENTUCKY RES FOUND.
 XX
 XX Schardl CL, Wang J;
 XX
 XX WPI; 2002-163205/21.
 XX

DR N-PSDB; ABK15520.
 XX
 PT New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 PT molecule) from fungi that are symbionts of commercially important
 PT grasses, useful to engineer ergot alkaloid-deficient symbionts.
 XX
 PS Claim 1; Col 11-14; 16pp; English.
 XX
 CC The invention describes an isolated dimethylallyltryptophan synthase
 CC (dmaW) nucleic acid (I). (I) is a dmaW sequence from epiphyte fungi
 CC Neotyphodium coenophialum symbiont of commercially significant grasses
 CC and is useful for expressing dmaW in a cell by recombinant techniques.
 CC (I) is also useful for: identifying endophytes e.g. from commercial
 CC plants used in forage, pasture, turf, land reclamation and soil
 CC conservation that contain or lack a dmaW gene and producing increased
 CC amount of ergot alkaloids involving expressing (I) in a host fungal cell
 CC so that the copy number of mRNA derived from transcription of the nucleic
 CC acid molecule is increased, allowing the host fungal cell to grow under
 CC appropriate growth conditions, thus causing increased production of ergot
 CC alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 CC or Acremonium or Epichloe that lack dmaW and therefore are unlikely to
 CC produce ergot alkaloids. The primers are useful for amplifying segments
 CC of dmaW from fungi in family Clavicipitaceae. (I) is also useful for
 CC identifying related sequences such as from Balansia, Balansioopsis,
 CC Echinodopsis, Atkinsonella, Myriogenospora, Neotyphodium, and
 CC Parepichloe, or natural or induced mutants. A knockout construct of (I)
 CC or antisense construct is useful for engineering ergot alkaloid-deficient
 CC fungal symbionts (endophytes of plants). This is the amino acid sequence
 CC of dimethylallyltryptophan synthase (dmaW) described in the method of the
 CC invention
 XX
 SQ Sequence 448 AA;
 Query Match 39.1%; Score 907.5; DB 5; Length 448;
 Best Local Similarity 41.3%; Pred. No. 2.6e-78;
 Matches 184; Conservative 78; Mismatches 157; Indels 27; Gaps 7;
 QY 7 AATLLPKFPYVLSQALNSKDHDKWYSTAPFATWAGAGYDVHQAQYKFLCIHREVII 66
 Db 4 AKTLHQEVYHTLSETFDANNDQRLMWHSTAPFEMKQLQTAHQLQAWISFSYKGGKPYMSV 63
 QY 67 PALGPYKPGQPMHWKSHLTFGLPFELSNYSKSLRFAFEPLGSLTGTKDDPFTQAI 126
 Db 64 PFLGVYPTSRGE-RWLSILTRYGTFPFLSNLSDSVRYTYEINATGSHLDPFTFAI 122
 QY 127 RPVLQDLKAMVPGDLDFWDFHTKALVVSBEAEARTLLDRDIEIPVKTQNKLAADLEPSG 186
 Db 123 WEALKKHIESQPGIDLEWDFSYFKQELTLDANESTYLHSONLVKREQIKTQNKALDLK-G 180
 QY 187 D-IVLKYIYPRIKSIATGTPKERMFDIAI-KAADKFGKVAATPLALEEPIAERAP--- 240
 Db 181 DXFVLKYIYPRIKSVATGKSVQELVFGSVKLAQKHKSIRPAFEMLEDIVQSRNFSST 240
 QY 241 -----TLL-GHFTSCDLVKPSESRIKVCWQERQDLASIEGIWTLNGRNDPDTLGLDA 294
 Db 241 DDSHNTLLSSRLSCDLISPTKSRVKTYLLERMVSLPAMEDLMTLGGRRDQSTIEGLEM 300
 QY 295 LRELWQLLPVTEGLCPNCFEYEPGTSPOQLPFIINFTLSPKALPEQIYEPAPGQND 354
 Db 301 IRELWGLNWSPLRAYPEPILPGAIPNEQLPSMANYTLHNDPIPEQVYFTVFGMND 360
 QY 355 KTAEGLATFFESRGWGLAKSPADLASYPDVLQTAHQLQAWISFSYKGGKPYMSV 414
 Db 361 MEVNTALTTFMRHESWDSMAKYKACLRSEFPDHDYALNYIHSYISFSYKKNKPYLSV 420
 QY 415 LHTFE-----AFSAAAEV 428
 Db 421 LHSFETGKVPVFPFEGTAFDCCRDL 446
 RESULT 3
 AAU76413
 ID AAU76413 standard; protein; 450 AA.

127 RVLQDLKAMVPGDLLEWFDHFTKALVSVSEEBARTLLDRDIEIPVFTQNKLAADLEPSG 186
 123 WEALKLIDSPGIDLQWFSYFKQELTLDANESTYLHSONLVKEQINTQNKALDLK-G 180
 187 D-IVLKYIYPRISATGTPKRLMEDAI-KAADKFGKATPLAILLEEFIAER--APT- 241
 181 DKFVLYIYIPKLSVATGKSVQELVFGSVKLAQKHSIRPAFEMLEDYVQSRNKVPTT 240
 242 -----LLGHFLSCDIVKPESEIRKIVYCMERQDLASIEGIWTLNGRNDPFTLGDGA 294
 241 DSHNTPLSRLSLSCDLVSPKSRVITYLLERVMVSLPAMEDLWTLGGRDQSTIEGLEM 300
 295 LRELWOLLPTVEGLCPPLNCFYBPGTSPQOLPFIINFTLSPKSALPEPOIYPAFGOND 354
 301 IRELWGLNMSFGLRAYPEPYLPGLAIPNQLPSMANYTLHNDPIPEPOVYFTVFGOND 360
 355 KTIAEGLATFFESRGWGLAKSPADLASIYDPVDLQTNANHLQAWISFSYKGGKPYMSVY 414
 361 MEVTNALTKEFMHEDMSMAKYKACLRSEFPFHNYEALNVIHSYISFSYRNNKPYSVY 420
 415 LHTFE 419
 421 LHSFE 425

RESULT 4
 RAO27492
 ID AAO27492 standard; protein; 880 AA.
 AC AAO27492;
 DT 06-NOV-2003 (first entry)
 DE Balsam pear (Momordica charantia) lipoxigenase isozyme 2 protein.
 KW Lipoxigenase; hydroperoxidation; polyunsaturated fatty acid; plant;
 fatty acid metabolite synthesis; signal molecule; growth regulation;
 development regulation; plant development; wound response;
 genetic mapping; hyperoxidation catalysis; Balsam pear; enzyme.
 OS Momordica charantia.
 PN US2003074693-A1.
 PD 17-APR-2003.
 PF 29-JAN-2002; 2002US-00059909.
 XX 10-FEB-1999; 99US-0119597P.
 PR 09-FEB-2000; 2000US-00501422.
 XX (CAHO//) CAHOON E B.
 PA (KINN//) KINNEY A J.
 PA (KLEI//) KLEIN T W.
 PA (LEEJ//) LEE J.
 PA (PEAR//) PEARLSTEIN R W.
 PA (RAFA//) RAFALSKI J A.
 PA (SHEN//) SHEN J B.
 PA (THOR//) THORPE C J.
 PA (TING//) TINGEV S V.
 PA (WENG//) WENG Z.
 XX Cahoon EB, Kinney AJ, Klein TM, Lee J, Pearlstein RW;
 PI Rafalski JA, Shen JB, Thorpe CU, Tingey SV, Weng Z;
 XX WPI; 2003-567325/53.
 DR N-PSDB; AAL57712.
 XX New isolated polynucleotides encoding plant lipoxigenases, useful in
 PT genetic mapping, particularly in catalyzing hyperoxidation of
 PR polyunsaturated fatty acids.
 XX
 PS Claim 19; Page 23-25; 36pp; English.

AAU76413;
 08-MAY-2002 (first entry)
 Dimethylallyltryptophan synthase (DmaW) version #2.
 Dimethylallyltryptophan synthase; dmaW; epiphyte fungus; symbiosis;
 endophyte; forage; pasture; turf; land reclamation; soil conservation;
 ergot alkaloid; Neotyphodium; Acremonium; Epichloe; Clavicipitaceae;
 Balansia; Balansioopsis; Echinoschloa; Atkinsonella; Myriogenospora;
 Neotyphodium; Paresichloe; knockout; antisense technology.
 Neotyphodium coenophialum.
 US6335188-B1.
 01-JAN-2002.
 03-MAR-2000; 2000US-00518657.
 22-MAR-1999; 99US-0125490P.
 (KENT) UNIV KENTUCKY RES FOUND.
 Schardl CL, Wang J;
 WPI; 2002-163205/21.
 N-PSDB; ABK15521.
 New isolated nucleic acid encoding dimethylallyltryptophan synthase (DmaW
 molecule) from fungi that are symbionts of commercially important
 grasses, useful to engineer ergot alkaloid-deficient symbionts.
 Claim 1; Col 15-18; 16pp; English.
 The invention describes an isolated dimethylallyltryptophan synthase
 (dmaW) nucleic acid (I). (I) is a dmaW sequence from epiphyte fungi
 Neotyphodium coenophialum symbiont of commercially significant grasses
 and is useful for expressing dmaW in a cell by recombinant techniques.
 (I) is also useful for: identifying endophytes e.g. from commercial
 plants used in forage, pasture, turf, land reclamation and soil
 conservation that contain or lack a dmaW gene and producing increased
 amount of ergot alkaloids involving expressing (I) in a host fungal cell
 so that the copy number of mRNA derived from transcription of the nucleic
 acid molecule is increased, allowing the host fungal cell to grow under
 appropriate growth conditions, thus causing increased production of ergot
 alkaloid. Probes derived from (I) are useful for identifying Neotyphodium
 or Acremonium or Epichloe that lack dmaW and therefore are unlikely to
 produce ergot alkaloids. The primers are useful for amplifying segments
 of dmaW from fungi in family Clavicipitaceae. (I) is also useful for
 identifying related sequences such as from Balansia, Balansioopsis,
 Echinoschloa, Atkinsonella, Myriogenospora, Neotyphodium, and
 Paresichloe, or natural or induced mutants. A knockout construct of (I)
 or antisense construct is useful for engineering ergot alkaloid-deficient
 fungal symbionts (endophytes of plants). This is the amino acid sequence
 of dimethylallyltryptophan synthase (dmaW) described in the method of the
 invention
 Sequence 450 AA;
 Query Match 39.0%; Score 904.5; DB 5; Length 450;
 Best Local Similarity 42.4%; Pred. No. 5.2e-78;
 Matches 180; Conservative 76; Mismatches 154; Indels 15; Gaps 6;
 7 AATLKPFPVLSQALNLSKDKTKWYSTAPMFATMAGAGVDVHAQYKFCIHEVII 66
 4 AKTLHGEVYTLSETDFDANNDRLLWHSTAPNFQKILQNTANYSIYAQYQHSIYKSHII 63
 67 PALGPYPEKQPMHWKSHLTFGLPPELSPFNYSKSLRFAFEPGLSLTGKDDPNTQAI 126
 64 PFLGVYPTSRGE-RWLSILTRYCTPPELSLNCSDSIVRYTYEPIAATGSHLDPFNTFAI 122

XX This invention relates to novel nucleotide sequences which encode
 CC proteins which have lipoxigenase activity. Lipoxigenases are membrane
 CC bound ubiquitous enzymes which catalyse the hydroperoxidation of
 CC polyunsaturated fatty acids in the first step of fatty acid metabolite
 CC synthesis. Products of this pathway are found as signal molecules
 CC involved in growth and development regulation. A knowledge of the amino
 CC acid sequence of lipoxigenases may allow the understanding of plant
 CC development and wound response. The polynucleotides, polypeptides and
 CC lipoxigenases of the invention may therefore be useful in genetic mapping
 CC and particularly for catalysing hydroperoxidation of polyunsaturated
 CC fatty acids. The present sequence is the amino acid sequence of the
 CC Balsam pear (*Momordica charantia*) lipoxigenase protein 2 of the invention
 XX
 SQ Sequence 880 AA;
 Query Match 4.5%; Score 103.5; DB 6; Length 880;
 Best Local Similarity 23.1%; Pred. No. 2.1;
 Matches 96; Conservative 46; Mismatches 138; Indels 135; Gaps 24;
 Qy 49 YDUHAQYKFLCIHR---EVIIPALG-----PYPEK---GQPMHWKSHLRFGLPFELSPN 97
 Db 221 YDV---YNDLCDPNGGPNLVRPILGSDQYFPRGRGTGRPPARKOHYESRLSDVMNLSN 277
 Qy 98 -YKSLRLFAPEPLGSLTGKDDPFNTQAIRPVLDLQKAMVPGLDLEWFDHFTKALVWSE 156
 Db 278 IYVPRDENFGHLKWADPLGNLTKVLST-SIQPGLESIFDSTPG-----EFDKP----- 324
 Qy 157 BEARTLDRDIETP--VFQTONKLAADLEPSGDIVLKYIYPRIKSIATGTPKERLMPDA 214
 Db 325 KEVDLDLPERGFPPIPLNFK---NLTEDLAPP-----EFTAEAPTLILGHFCLSDLVKSESRI 356
 Qy 215 IKAAD--KFGKVATPLAILE-----EFTAEAPTLILGHFCLSDLVKSESRI 259
 Db 357 FLRSDGRFLUKYFPQVQKNGKLGWRTDEFAEMAGVNPPLII-----REKL 403
 Qy 260 KYCYMERQLDLASIEGIWTLNGRRNDPET-----LDGL---DALRE-LWOLLPVTEGL 308
 Db 404 EVFPLSLKLP-----HVVGNQNSTMTTEEQIKHGLDGLTVDEAIKENKLYILDHHDAL 456
 Qy 309 CPLPNCYEGCTSPQQLPIL---NFTLSP---KSALPEPQ-----IYFPAFGQN 353
 Db 457 MPYLRRINSTKTYATRTLLFLKDDSTLKLALIELSLPHQGDHGAISKLYFFAAGR- 515
 Qy 354 DKTIAGLATFFSRGNGGGLAKSPADLASYYDPVD---LQTAHLQAWISFSYK 405
 Db 516 -----VESAIW-QLAKEYAVVNDSGYHQLNSHLTHAVLEPFIITHR 558
 RESULT 5
 AAM50327
 ID AAM50327 standard; protein; 521 AA.
 XX
 AC AAM50327;
 XX
 XX 04-FEB-2002 (first entry)
 DT
 XX Human nucleotide binding site protein NBS-4.
 DE
 XX NBS-4; nucleotide binding site; human; antiinflammatory; antiapoptotic;
 XX cytotatic; antimicrobial; anti-HIV; antiparkinsonian; antianemic;
 XX neuroprotective; nootropic; cardiatic; cerebroprotective; antiarthritic;
 XX antidiabetic; immunosuppressive; thyronimetic; antibacterial;
 XX tuberculostatic; virucide; signal transduction; vaccine; therapy;
 XX diagnosis.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Modified-site 11..14
 FT Modified-site /note= "Asn is N-glycosylated"
 FT Modified-site 23..26
 FT Bertin J;

FT Modified-site 24..27 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 24..26 /note= "O-phosphorylated by protein kinase C"
 FT Modified-site 39..42 /note= "Asn is N-glycosylated"
 FT Domain 42..521 /label= Nucleotide-binding_site
 FT Domain 47..62 /label= Kinase-1a_domain
 FT Binding-site 52..59 /note= "P-loop"
 FT Modified-site 55..60 /note= "ATP/GTP-binding site motif A (P loop)"
 FT Modified-site 97..100 /note= "N-myristoylated"
 FT Domain 116..132 /label= Kinase-2a_domain
 FT Modified-site 118..121 /note= "Walker B Box"
 FT Modified-site 138..141 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 142..145 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 146..151 /note= "O-phosphorylated by casein kinase II"
 FT Domain 174..185 /label= Kinase-3a_domain
 FT Modified-site 179..181 /note= "N-myristoylated"
 FT Modified-site 204..207 /note= "O-phosphorylated by protein kinase C"
 FT Modified-site 220..223 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 232..235 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 282..285 /note= "Asn is N-glycosylated"
 FT Modified-site 295..300 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 312..315 /note= "N-myristoylated"
 FT Modified-site 320..323 /note= "Asn is N-glycosylated"
 FT Modified-site 332..335 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 346..351 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 355..358 /note= "N-myristoylated"
 FT Modified-site 377..379 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site 426..428 /note= "O-phosphorylated by protein kinase C"
 FT Modified-site 462..465 /note= "O-phosphorylated by protein kinase C"
 FT Modified-site 468..470 /note= "O-phosphorylated by casein kinase II"
 FT Modified-site /note= "O-phosphorylated by protein kinase C"
 WO200183753-A2.
 08-NOV-2001.
 03-MAY-2001; 2001WO-US014283.
 03-MAY-2000; 2000US-0201464P.
 (MILL-) MILLENNIUM PHARM INC.
 Bertin J;

DR WPI; 2002-041495/05.
DR N-PSDB; AA170682.
XX
XX Novel isolated polypeptide of nucleotide binding site useful as a vaccine
PT for preventing or treating diseases e.g. cancer, stroke, Alzheimer's
PT disease, Parkinson's disease, myocardial infarction, Crohn's disease.
XX
XX Claim 12; Fig 9A-C; 149pp; English.
XX
XX The present sequence is that of novel human NBS-4, as deduced from a
CC partial cDNA clone (see AA170682). The sequence is predicted not to be
CC full length. The human NBS-4 protein has a nucleotide binding site, which
CC is present in a number of proteins that transmit signals which activate
CC apoptotic and inflammatory pathways in response to stress and other
CC stimuli. The predicted full-length protein should also have a leucine-
CC rich repeat domain (LRR), which is also present in proteins involved in
CC apoptotic and inflammatory pathways, and a pyrin domain. The invention
CC provides NBS-2, NBS-3, NBS-4 and NBS-5 nucleic acids (see AA170680-85)
CC and polypeptides (see AAM50325-30), as well as modulators of NBS-2, NBS-
CC 3, NBS-4 and NBS-5 activity or expression. These are expected to be
CC useful in the modulation of stress-related, apoptotic and inflammatory
CC responses. The nucleic acids and polypeptides are also expected to be
CC useful in diagnosis of such disorders and in screening assays used to
CC identify modulator compounds. The inflammatory and immune disorders
CC include, but are not limited to, chronic inflammatory diseases and
CC disorders, such as Crohn's disease, reactive arthritis, including Lyme
CC disease, insulin-dependent diabetes, organ-specific autoimmunity,
CC including multiple sclerosis, Hashimoto's thyroiditis and Grave's
CC disease, contact dermatitis, psoriasis, graft rejection, graft versus
CC host disease, sarcoidosis, atopic conditions, such as asthma and allergy,
CC including allergic rhinitis, gastrointestinal allergies, including food
CC allergies, eosinophilia, conjunctivitis, glomerular nephritis, and
CC infections including helminthic (e.g. leishmaniasis), viral (e.g. HIV),
CC and bacterial (e.g. tuberculosis and lepromatous leprosy). Apoptotic
CC pathway disorders include cancer, infectious diseases and autoimmune
CC disorders such as systemic lupus erythematosus, immune-mediated
CC glomerulonephritis and arthritis, as well as neurological diseases
CC including Alzheimer's disease, Parkinson's disease, amyotrophic lateral
CC sclerosis, retinitis pigmentosa and Huntington's disease, and
CC haematological diseases including aplastic anaemia, chronic neutropaenia
CC and myelodysplastic syndromes
XX
XX Sequence 521 AA;
SQ
Query Match 4.4%; Score 102; DB 5; Length 521;
Best Local Similarity 22.3%; Pred. No. 1.4; 138; Indels 94; Gaps 18;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPFELSFNYSKGLRFAFELPGLSLGTGK-DDPFNTQAIRPVLDLQK 134
Db 67 LHWANGVLFOQRFSYVYLSCHKIRYKETTFAELISLDWDPDFAPIEFMSQP--EKLL 124
QY 135 AMVPGDLEWDFHTKALVVSSEBEARTLDRD-----IEIPVK-TQNKLAADLEPSGD 187
Db 125 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTKILHLKELVPLAT 175
QY 188 --IVLKTIVYPRKSI-----ATGTPKERMFDIAKADKFGKVPATPLAILEE--- 233
Db 176 LLITIKTWFRDLKASLVNPFVQITGTDGLDRVYFWRHFDSDSEVEKILQQLRKNETL 235
QY 234 FIAERAPTLGLHFLSC-----DLVKPSEIRIKYVC-----MERQLDIA----- 271
Db 236 FHSCSAPMVCWTCVCSLQKQKRYVDLQISITQTTSLYAFNLFSTAEVDLADDSWPG 295
QY 272 -----STEGITWTLNGRNDPET-LDGI-----DALRELWOLLFVTE-GLCPL----- 311
Db 296 QWRALCSLAEGLWMNFTNKEDTEIEGLEVFIDLSLVEFNILQKINDCGGCTTHLS 355
QY 312 -----PNCFYBPGTSQQLPFIINFTLSPKSALPEPQIVPPAFGQNDKTI 357
Db 356 QBFPAAMSVLEEFREPPHSTKQPB-MWMLQHVLLDKEAYWTFVWLPF-PGLLNKNI 413
QY 358 AEGL 361

Db 414 AREL 417
Py Pyrin domain containing protein NALP13/Py17.
XX
XX 20-AUG-2002 (first entry)
XX
XX Pyrin domain; PYD domain; antiinflammatory; antiparkinsonian;
antiarteriosclerotic; antipsoriatic; antibacterial; virucide;
neuroprotective; antiarthritic; antirheumatic; antiasthmatic;
KW neurotropic; osteopathic; nontropic; intracellular signal transduction;
KW inflammation; Alzheimer's disease; infection; psoriasis; asthma;
KW arteriosclerosis; multiple sclerosis; rheumatoid arthritis; sarcoidosis;
KW osteoarthritis; glomerulonephritis.
XX
XX Unidentified.
XX
XX WO200240668-A2.
XX
XX 23-MAY-2002.
XX
XX 30-OCT-2001; 2001WO-EP012545.
XX
XX 15-NOV-2000; 2000DE-01056687.
XX
XX 30-NOV-2000; 2000DE-01059595.
XX
XX (APOT-) APOTECH RES & DEV LTD.
XX
XX Tschopp J, Martinon F;
XX
XX WPI; 2002-427093/45.
XX
XX N-PSDB; AA147140.
XX
XX New DNA encoding protein with pyrin domain, useful for treating diseases
PT involving impaired signal transduction, particularly inflammation, also
PT proteins and antibodies.
XX
XX Claim 5; Fig 1; 116pp; German.
XX
XX The present invention relates the DNA and their encoded proteins, where
CC the proteins contain at least one PYD (pyrin) domain. These can be used
CC to treat diseases associated with impaired intracellular signal
CC transduction, particularly inflammation such as psoriasis,
CC arteriosclerosis, bacterial or viral infections (particularly meningitis
CC and pneumonia), multiple sclerosis, rheumatoid arthritis, asthma,
CC sarcoidosis, glomerulonephritis and osteoarthritis, and also Alzheimer's
CC and Parkinson's diseases. The present sequence is a protein of the
CC invention
XX
XX Sequence 606 AA;
SQ
Query Match 4.4%; Score 102; DB 5; Length 606;
Best Local Similarity 22.3%; Pred. No. 1.7; 139; Indels 94; Gaps 18;
Matches 81; Conservative 51; Mismatches 139; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPFELSFNYSKGLRFAFELPGLSLGTGK-DDPFNTQAIRPVLDLQK 134
Db 232 LHWANGVLFOQRFSYVYLSCHKIRYKETTFAELISLDWDPDFAPIEFMSQP--EKLL 289
QY 135 AMVPGDLEWDFHTKALVVSSEBEARTLDRD-----IEIPVK-TQNKLAADLEPSGD 187
Db 290 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTKILHLKELVPLAT 340
QY 188 --IVLKTIVYPRKSI-----ATGTPKERMFDIAKADKFGKVPATPLAILEE--- 233
Db 341 LLITIKTWFRDLKASLVNPFVQITGTDGLDRVYFWRHFDSDSEVEKILQQLRKNETL 400

QY 234 FIABRAPTLGHFLSC-----DLVKPSBSRIKVC-----MERQDLIA----- 271
 Db 401 FHSCSAPMVCWTGSCLPKRVYDIQSITQTTSILYAFFNLSFSTAEVDLADDSWPG 460
 QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWOLLPVE-GLCPL----- 311
 Db 461 QWRALCSLAIEGLSMNFTNKEDTIEGLEVPFIDSLYEFNLQINDCGGTTTHLS 520
 QY 312 -----PNCFYEPGTSPOQLPFIINFTLSPKSLPEPQIYFPAGQNDKTI 357
 Db 521 FQEFFAAMSVLEBEPFPFPHSTKQOE-MQMLLQHVLLDKAEYWTVPVILF-PFGLLNKNI 578
 QY 358 AEG 361
 Db 579 AREL 582

RESULT 7
 ID ABO07113 standard; protein; 635 AA.
 AC ABO07113;
 XX 13-AUG-2003 (first entry)
 DE Novel human protein NOV2a.

XX NOVX; autoimmune disease; allergy; Alzheimer's disease; stroke;
 XX Parkinson's disease; Huntington's disease; multiple sclerosis; addiction;
 XX anxiety; pain; diabetes; glomerulonephritis; obesity;
 XX systemic lupus erythematosus; asthma; scleroderma; pancreatitis;
 XX graft versus host disease; ulcer; anaemia; cancer; trauma; infection;
 XX cardiomyopathy; atherosclerosis; hypertension; AIDS; Crohn's disease;
 XX acquired immunodeficiency syndrome; chromosomal mapping; tissue typing;
 XX forensic biology; predictive medicine; gene therapy; human.

OS Homo sapiens.
 XX
 XX WO200298900-A2.
 XX
 XX 12-DEC-2002.
 XX
 XX 04-JUN-2002; 2002WO-US017558.
 XX
 XX 04-JUN-2001; 2001US-0295607P.
 XX 04-JUN-2001; 2001US-0295661P.
 XX 06-JUN-2001; 2001US-0296404P.
 XX 06-JUN-2001; 2001US-0296418P.
 XX 07-JUN-2001; 2001US-0296757P.
 XX 11-JUN-2001; 2001US-0297414P.
 XX 12-JUN-2001; 2001US-0297567P.
 XX 15-JUN-2001; 2001US-0298528P.
 XX 18-JUN-2001; 2001US-0299133P.
 XX 19-JUN-2001; 2001US-0299230P.
 XX 21-JUN-2001; 2001US-0299499P.
 XX 22-JUN-2001; 2001US-0300177P.
 XX 26-JUN-2001; 2001US-0300883P.
 XX 28-JUN-2001; 2001US-0301530P.
 XX 28-JUN-2001; 2001US-0301550P.
 XX 03-JUL-2001; 2001US-0302951P.
 XX 12-SEP-2001; 2001US-0318727P.
 XX 27-SEP-2001; 2001US-0325685P.
 XX 22-FEB-2002; 2001US-0358814P.
 XX 03-JUN-2002; 2002US-00161927.

(CURA-) CURAGEN CORP.
 XX
 XX Zerhusen BD, Kekuda R, Spytek KA, Shenoy SG, Miller CE, Hialt T;
 XX Gerlach VL, Baumgartner JC, Guo X, Gangolli EA, Vernet CAM;
 XX Padigar M, Li L, Pena CE, Gorman L, Anderson DW, Edinger SR;
 XX Paturajan M, Stone DJ;

DR WPI; 2003-140585/13.
 DR N-PSDB; ACD13187.
 XX
 PT Novel isolated NOVX polypeptide useful treating or preventing disorders
 PT or syndromes such as autoimmune disease, allergies, Alzheimer's disease,
 PT stroke, Parkinson's disease, Huntington's disease or multiple sclerosis.
 XX
 XX Claim 1; Page 87; 408pp; English.
 PS
 XX The invention describes an isolated NOVX polypeptide (I) comprising a
 CC sequence selected from a sequence (S1) of 1121, 635, 299, 1720, 176, 583,
 CC 214, 395, 1098, 134, 427, 1333, 407, 806, 804, 1253, 382, 1045, 284, 496,
 CC 506, 759, 390, 133, 215, 240, 1069, 116, 439, 1138, 477, 316, 269, 219,
 CC 305, 406, 460, 365, 380, 829 or 326 amino acids fully defined in the
 CC specification, and the mature form of Si. (I) is useful for treating or
 CC preventing a pathology associated with (I) in a subject, preferably
 CC human, or for identifying an agent that binds to (I), where the agent is
 CC a cellular receptor or a downstream effector. (I), a polynucleotide (II)
 CC encoding (I) or an anti-(I)-antibody (V) is useful treating or preventing
 CC disorders or syndromes such as autoimmune disease, allergies, Alzheimer's
 CC disease, stroke, Parkinson's disease, Huntington's disease, multiple
 CC sclerosis, addiction, anxiety, pain, diabetes, glomerulonephritis,
 CC systemic lupus erythematosus, asthma, scleroderma, graft versus host
 CC disease, pancreatitis, obesity, ulcers, anaemia, cancer, trauma, viral,
 CC bacterial or parasitic infections, cardiomyopathy, atherosclerosis,
 CC hypertension, acquired immunodeficiency syndrome (AIDS) or Crohn's
 CC disease. (I), (II) or (V) is useful in screening assays, detection assays
 CC (e.g., chromosomal mapping, tissue typing, forensic biology), predictive
 CC medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical
 CC trials and pharmacogenomic), and in methods of treatment (e.g.,
 CC therapeutic and prophylactic). (II) is useful in gene therapy, to express
 CC (I), to detect NOVX mRNA or a genetic lesion in a NOVX gene, and to
 CC modulate NOVX activity. This is the amino acid sequence of a novel human
 CC NOV protein
 XX
 XX Sequence 635 AA;
 XX
 XX Query Match 4.4%; Score 102; DB 6; Length 635;
 XX Best Local Similarity 22.3%; Pred. No. 1.8;
 XX Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
 QY 79 MHKSHL---TRGLPELSFNYSKSLRPAFPLSLGTGK-DDPNTQAIRPVLDLX 134
 Db 67 LHWANGVLFQQRFSYVFLSKIRYKETTFAELISLDWDFDPAIEFMSQF--EKLL 124
 QY 135 AMVEGLDLEWDFHTKALVSEEARLTLDRD-----HEIPVEK-TQNKLAADLPSPGD 187
 Db 125 FIIDGFE-----EIIISERSSESLDDGSPECTWYQELPVTKILHSLKXELPLAT 175
 QY 188 --IVLKYIYPRKSI-----AGTPKERMFAKAADKFKGVAPLAIEE---- 233
 Db 176 LLITIKTFVFRDLKASLWNPFCVQITGTDGLRVVFMRFHDDSSSEVEKILQLRKNETL 235
 QY 234 FIABRAPTLGHFLSC-----DLVKPSBSRIKVC-----MERQDLIA----- 271
 Db 236 FHSCSAPMVCWTGSCLPKRVYDIQSITQTTSILYAFFNLSFSTAEVDLADDSWPG 295
 QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWOLLPVE-GLCPL----- 311
 Db 296 QWRALCSLAIEGLSMNFTNKEDTIEGLEVPFIDSLYEFNLQINDCGGTTTHLS 355
 QY 312 -----PNCFYEPGTSPOQLPFIINFTLSPKSLPEPQIYFPAGQNDKTI 357
 Db 356 FQEFFAAMSVLEBEPFPFPHSTKQOE-MQMLLQHVLLDKAEYWTVPVILF-PFGLLNKNI 413
 QY 358 AEG 361
 Db 414 AREL 417
 XX
 XX RESULT 8
 XX AAB11500

QY AAB11500; 366 ESRGWGLAKSPADLASYPVDLQTANHLQAWISFYKGGKPKYMSVYLHT---FEAFS 422
 AC 13-MAR-2001 (first entry) 519 ---W-QAKAY-----VTVDVGYHQLISHWLTHTAVLEPFV 551
 DT 519 ---W-QAKAY-----VTVDVGYHQLISHWLTHTAVLEPFV 551
 XX C. sativus LOX protein.
 DE LOX; lipoxigenase; cucumber; plant; 6-hydroperoxy-gamma-linolenic acid;
 KW 9-hydroperoxy-gamma-linolenic acid; 6,9-hydroperoxy-gamma-linolenic acid;
 KW gamma-linolenic acid.
 XX Cucumis sativus.
 OS DE19914464-A1.
 XX DE19914464-A1.
 XX 05-OCT-2000.
 XX 30-MAR-1999; 99DE-01014464.
 XX 30-MAR-1999; 99DE-01014464.
 XX (TPBP-) IPB INST PFLANZENBIOCHEMIE.
 XX Feussner I, Hornung E;
 XX WPI; 2001-103874/12.
 XX Preparation of plant lipoxigenase with altered position specificity
 PT comprises replacing at least one amino acid in the wild type lipoxigenase
 PT amino acid sequence.
 XX Disclosure; Fig 5; 14pp; German.
 XX This invention describes a novel preparation of plant lipoxigenase (I)
 CC with altered position specificity which comprises replacing at least 1
 CC amino acid in the wild type lipoxigenase amino acid sequence. The method
 CC also describes (1) (I) prepared by the method above; (2) nucleic acid
 CC encoding (I); (3) a vector comprising the nucleic acid of (2); (4) a cell
 CC comprising the vector of (3); (5) a plant or plant part comprising a host
 CC cell as in (4); (6) preparation of 6-, 9- and/or 6,9-hydroperoxy-gamma-
 CC linolenic acid comprising reacting gamma-linolenic acid with (I); and (7)
 CC a gamma-linolenic acid derivative comprising a hydroperoxy group or a
 CC hydroxy group at position 6
 XX Sequence 878 AA;
 SQ Query Match 4.3%; Score 101; DB 4; Length 878;
 Best Local Similarity 21.7%; Pred. No. 3.7;
 Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;
 QY 61 HREVIIPALG-----PYPEK---GQPMHWKSHLTRFGLPELSFNYSKSLRFAPEPLGS 112
 DB 235 HR-----PILGGTTEYPYPRRGRTGRPSRDH-----NYESRL-----SPIMS 273
 QY 113 LT--GTKDDPFN-----TQAIRPVLDLKAMVPGLDLEWFDHFHTKALVVS 155
 DB 274 LDIVPKDENFGHLKMSDFGLYTLKALSISIKPGLQSFVTP-----NEFDNF----- 322
 QY 156 EBEARTLLDRDIEIP--VFQTKNKLADLEPSGDIVLKYIYPRIKSIATGTPKRLMFD 213
 DB 323 -KEVDNLFERGFPIPFNAFKT---LTEDLTP-----ELFKALVRNDGKFLKFP 367
 QY 214 AIKAADKFGKV--ATPLAILEFTAEAPTLLGHFLSCDLVKPSSESRKIVYCMERQDLA 271
 DB 368 TPVVU-KONKIGSTDEFAREMLAGNPILLI-----RLEAFPTSKLDP- 412
 QY 272 SIEGIWTLGRNDPDT-----LDGLDALRELWQ-----LLPVTEGLCPNCFVEPGT 320
 DB 413 -----NVYGNQNSTIEEHIKGLDGLTVDEAMQNRLYIVDFHDALPFLYTRMNTAT 466
 QY 321 SPQQLPFI--NFTLSP-----KSALEPQO-----LYEFAFGQNDKTIAEGLATPF 365
 DB 467 KTVATRTLLKDDGTGLTKPLVIELALPHPOQDGLGALSILKYFFAENGUVQKSI----- 518

QY AAB11500; 366 ESRGWGLAKSPADLASYPVDLQTANHLQAWISFYKGGKPKYMSVYLHT---FEAFS 422
 AC 13-MAR-2001 (first entry) 519 ---W-QAKAY-----VTVDVGYHQLISHWLTHTAVLEPFV 551
 DT 519 ---W-QAKAY-----VTVDVGYHQLISHWLTHTAVLEPFV 551
 XX C. sativus LOX protein.
 DE LOX; lipoxigenase; cucumber; plant; 6-hydroperoxy-gamma-linolenic acid;
 KW 9-hydroperoxy-gamma-linolenic acid; 6,9-hydroperoxy-gamma-linolenic acid;
 KW gamma-linolenic acid.
 XX Cucumis sativus.
 OS DE19914464-A1.
 XX DE19914464-A1.
 XX 05-OCT-2000.
 XX 30-MAR-1999; 99DE-01014464.
 XX 30-MAR-1999; 99DE-01014464.
 XX (TPBP-) IPB INST PFLANZENBIOCHEMIE.
 XX Feussner I, Hornung E;
 XX WPI; 2001-103874/12.
 XX Preparation of plant lipoxigenase with altered position specificity
 PT comprises replacing at least one amino acid in the wild type lipoxigenase
 PT amino acid sequence.
 XX Disclosure; Fig 5; 14pp; German.
 XX This invention describes a novel preparation of plant lipoxigenase (I)
 CC with altered position specificity which comprises replacing at least 1
 CC amino acid in the wild type lipoxigenase amino acid sequence. The method
 CC also describes (1) (I) prepared by the method above; (2) nucleic acid
 CC encoding (I); (3) a vector comprising the nucleic acid of (2); (4) a cell
 CC comprising the vector of (3); (5) a plant or plant part comprising a host
 CC cell as in (4); (6) preparation of 6-, 9- and/or 6,9-hydroperoxy-gamma-
 CC linolenic acid comprising reacting gamma-linolenic acid with (I); and (7)
 CC a gamma-linolenic acid derivative comprising a hydroperoxy group or a
 CC hydroxy group at position 6
 XX Sequence 878 AA;
 SQ Query Match 4.3%; Score 101; DB 4; Length 878;
 Best Local Similarity 21.7%; Pred. No. 3.7;
 Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;
 QY 61 HREVIIPALG-----PYPEK---GQPMHWKSHLTRFGLPELSFNYSKSLRFAPEPLGS 112
 DB 235 HR-----PILGGTTEYPYPRRGRTGRPSRDH-----NYESRL-----SPIMS 273
 QY 113 LT--GTKDDPFN-----TQAIRPVLDLKAMVPGLDLEWFDHFHTKALVVS 155
 DB 274 LDIVPKDENFGHLKMSDFGLYTLKALSISIKPGLQSFVTP-----NEFDNF----- 322
 QY 156 EBEARTLLDRDIEIP--VFQTKNKLADLEPSGDIVLKYIYPRIKSIATGTPKRLMFD 213
 DB 323 -KEVDNLFERGFPIPFNAFKT---LTEDLTP-----ELFKALVRNDGKFLKFP 367
 QY 214 AIKAADKFGKV--ATPLAILEFTAEAPTLLGHFLSCDLVKPSSESRKIVYCMERQDLA 271
 DB 368 TPVVU-KONKIGSTDEFAREMLAGNPILLI-----RLEAFPTSKLDP- 412
 QY 272 SIEGIWTLGRNDPDT-----LDGLDALRELWQ-----LLPVTEGLCPNCFVEPGT 320
 DB 413 -----NVYGNQNSTIEEHIKGLDGLTVDEAMQNRLYIVDFHDALPFLYTRMNTAT 466
 QY 321 SPQQLPFI--NFTLSP-----KSALEPQO-----LYEFAFGQNDKTIAEGLATPF 365
 DB 467 KTVATRTLLKDDGTGLTKPLVIELALPHPOQDGLGALSILKYFFAENGUVQKSI----- 518

368 TPVV-KDKLGWSTDEFAEMLAGNPLLI-----RLEAFPTSKLDP- 412
 272 SIEGWTNGRNDPET-----LDGLDALRELWQ-----LLPVTEGLCLPNCFYEPGT 320
 413 -----NVYGNQNSTEIEHKHGLDGLTVDEAMQNRLVIVDFHDALMFLYRMTATST 466
 321 SPQQLPFI--NFTLSP--KSALEPQ-----IYFAFGQNDKTIAGLATFF 365
 467 KYATRTLLKDDGTGLKPLVIELPHPODQGLAISKLYFAENGVSQSI----- 518
 366 ESRGWGLAKSYPADLASYPVDLQTAHLOAWISFYKGYKPYMSVLIHT---PEAFS 422
 519 ----W-QLAKAY-----VTVDVGYHQLISHLHTHAVLEPFV 551
 423 AAA-QEVAMCHDGH 435
 552 IATHRQLSVLPHI 565

RESULT 10
 ID ADE54304 standard; protein; 503 AA.
 AC ADE54304;
 DT 29-JAN-2004 (first entry)
 DE Rat Protein BAA20354, SEQ ID NO 107.
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 OS Rattus norvegicus.
 PN WO2003016475-A2.
 PD 27-FEB-2003.
 PF 14-AUG-2002; 2002WO-US025765.
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX (GEMO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; BAA20354.
 XX

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.
 Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 503 AA;

Query Match 4.3%; Score 99.5; DB 7; Length 503;
 Best Local Similarity 19.3%; Pred. No. 2.3;
 Matches 92; Conservative 62; Mismatches 190; Indels 133; Gaps 19;

QY 39 MFATMMAGAGYDVHAQYKF-LCIHREVIIPALGPYPEK-GQPMHWKSHLRFGLPFEELS 95
 Db 26 LLSTLLIACAFSLSLVYLFRLAVGHVQLPAGAKSPPIYSPFPLGHAFAGKSPLEFL 85
 QY 96 FN-YSKSLRFAPEPLGS-----LTKDPPFNTQAI-----RPV----- 129
 Db 86 ENAYEKYGPVPSFTWVGKTFYLLGSDAALLFNSKNEDLNAERVYGRLLTPVPFGKGVAY 145
 QY 130 -----LQDLKAMVPGDLEWFDHFTKALVVSSEAR----- 160
 Db 146 DVPNAVLEQKILKSGNLIA---HFQYVSIIEKEAKYFKSWGESGERNVFEALSELI 202
 QY 161 -----TLLDRDIBIPVFKTQNKLAADLE-----PS-----GDIVLKT 192
 Db 203 ILTASHCLHGKEIRSQLNEKVAQYADLDGFSHAALLPGWLPSPFRDRRAHREIKN 262
 QY 193 YIYPRI-KSATGTPKERMFDALKAADKFGKVATP-----LAILBEPTAEARPTL 242
 Db 263 IFYKAIQKRRLSKPEADILQTLDDSTYKDGRLPTDDEIAGMLIGLLLAGQHTSSTTSAW 322
 QY 243 LGHFLSCDLVKPSESRI---KVVC-----MERQLDLASIE-----GIWTL 279
 Db 323 MGFFLARD--KPLQDKCYLBQKTVCGEDLPPLYEQKDLNLDRCIKETLRLRPPIWTM 380
 QY 280 NGRNDPETLDGLDALRELWQLLPVTEGLCPNCFYEPGTSPQEQLPFIINFTLSPKSA 339
 Db 381 MRMAKTPQTIVAG-----YTTPPGHQVCVSPVTNQRKDSWVERLDENPDYLDQNP 432
 QY 340 LPEPQIYPPAFGQNDKTIAGSLATFFESRCGWSGLAKSYPADLAS--YVPDVLQTAH 395
 Db 433 SGEKFAIVPFGAGRHRCIGENFAIVQKTIWSTMLRLYEFDLNGIPPSVNYTTLIH 489

RESULT 11
 ADE54292
 ID ADE54292 standard; protein; 503 AA.
 XX
 AC ADE54292;
 XX
 DT 29-JAN-2004 (first entry)
 DE Rat Protein BAA20354, SEQ ID NO 95.
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 OS Rattus norvegicus.
 XX
 PN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 PF 14-AUG-2002; 2002WO-US025765.
 XX

RESULT 14
ADE62266
ID ADE62266 standard; protein; 503 AA.
XX AC ADE62266;
XX DT 29-JAN-2004 (first entry)
XX DE Rat Protein Q64654, SEQ ID NO 8195.
XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX KW Chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX OS Rattus norvegicus.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GHEO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX DR GENBANK; Q64654.
XX PT New composition comprising two or more isolated polypeptides, useful for
XX PT preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat
or human polynucleotides or a polynucleotide which represents a fragment,
derivative or allelic variation of the nucleic acid sequence. Also
claimed are a vector comprising the novel polynucleotide, a host cell
comprising the vector, a method for identifying a nucleotide sequence
which is differentially regulated in an animal subjected to pain and a
kit to perform the method, an array, a method for identifying an agent
that increases or decreases the expression of the polynucleotide sequence
that is differentially expressed in neuronal tissue of a first animal
subjected to pain, a method for identifying a compound which regulates
the expression of a polynucleotide sequence which is differentially
expressed in an animal subjected to pain, a method for identifying a
compound that regulates the activity of one or more of the
polynucleotides, a method for producing a pharmaceutical composition, a
method for identifying a compound or small molecule that regulates the
activity in an animal of one or more of the polypeptides given in the
specification, a method for identifying a compound useful in treating
pain and a pharmaceutical composition comprising the one or more
polypeptides or their antibodies. The polynucleotide or the compound that
modulates its activity is useful for preparing a medicament for treating
pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
therapy). The sequence presented is a rat protein (shown in Table 2 of
the specification) which is differentially expressed during pain. Note:
The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
ftp.wipo.int/pub/published_pat_sequences.
XX SQ Sequence 503 AA;
Query Match 4.3%; Score 99.5; DB 7; Length 503;
Best Local Similarity 19.3%; Pred. No. 2.3;
Matches 92; Conservative 62; Mismatches 190; Indels 133; Gaps 19;
39 MPATMAGAGYDVHAQYKP-LCIHREVILFALGPYBEK-QOPMHKSHLTFPG-LPELS 95
26 LLSTLLIACAPFTLSLVFLFELAVGHMVOLPAGAKSPPIYISPIFLCHALAFKSPLEFL 85
96 FN-YSKSLRLFAFEPLGS-----LTGTDDPFTQAI-----RPV----- 129
86 ENAYEKYGPVFTMVGKFTTYLLGSDAAALLFNSKXNEDLNABEVYGRLLTPVFGKGVAY 145
130 -----LQDLKAMVPOLDLEWFDHFTKALVNSEEAR----- 160
146 DVPNAVLEQKIKLKSGLNIA---HFQYVSIIEKEAKEYFKSGWGESGERNVFEALSELI 202
161 -----TLLDRDRIPIPVFKTONKLAADLE-----PS-----GDIVLKT 192
203 ILTASHCLHGKEIRSQLNEKVAQLYADLDGFSHAALLPGWLPLPSFRDRRAHREIKN 262
193 YIYPRI-KSIATGTPKERLMFMDAIAADKFGKVATP-----LAILLEPTAERAPTL 242
263 IFYKAIQKRRLSKPEADILQTLIDSTYKDGRLTDDDEIAGMLIGLLLAGQHTSSTTSAW 322
243 LGHFLSCDLVKPSESRI---KYVC-----MERQLDLASIE--- 279
323 MGFFLRD--KPLQDKCYLEQKTVCGEDLPLLYEQKLDNLDRCIKETLRLEPPMTM 380
280 NGRNDPPTLDGLDALRELWQLLPVTEGLCPNCFYEPGTSPOEQLPFFIINFTLSPKSA 339
381 MRVAKTPTQTVAG-----YTIPPGHQVCVSPVTNQRLKDSWVERLDNPDRLQDNPA 432
340 LPEPQIYPPAFGQNDKTIARGLATFFESRGGWGLAKSPADLAS-YYPDVLDQTAH 395
433 SGKEFAYVPPFAGRHRCIGENFAYVQIKTIWSTMLRLYEFDLNGYFPFSVNYTTMIH 489
RESULT 15
ADE54300
ID ADE54300 standard; protein; 503 AA.
XX AC ADE54300;
XX DT 29-JAN-2004 (first entry)
XX DE Rat Protein BAA20354, SEQ ID NO 103.
XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX KW Chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX OS Rattus norvegicus.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GHEO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX DR GENBANK; BAA20354.
XX PT New composition comprising two or more isolated polypeptides, useful for
XX PT preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat

	Query Match	100.0%;	Score 2322;	DB 4;	Length 437;
	Best Local Similarity	100.0%;	Pred. No. 3.1e-235;		
	Matches 437;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MEISKKAATLLPKPFYVLSQALNSLNKDTHKWWYSTAPMEATMAGAGYDVHQAQYKFLCI	60		
Db	1	MEISKKAATLLPKPFYVLSQALNSLNKDTHKWWYSTAPMEATMAGAGYDVHQAQYKFLCI	60		
Qy	61	HREVIIPALGPYPEKGOPMEHWKSHLTFGLPFELSFNYSKSLRFAPEPLGSLTGTGKDDP	120		
Db	61	HREVIIPALGPYPEKGOPMEHWKSHLTFGLPFELSFNYSKSLRFAPEPLGSLTGTGKDDP	120		
Qy	121	FNTOAIRPVLQDIKAMVPGLDLSWDFHTFKALVVSBEEARLLDRDIEIIPVFTQNKLA	180		
Db	121	FNTOAIRPVLQDIKAMVPGLDLSWDFHTFKALVVSBEEARLLDRDIEIIPVFTQNKLA	180		
Qy	181	DLBPSGDIVLTKVIYPRIKSIATGTPKXERLMFDAIKAADFGKVAFTPLAILEEFIABRAP	240		

Db 181 DLEPSSGIVLTKYIYPRIKSIATGTPKERMFDALKAADKFGKVATPLAILEEFIAERAP 240
Qy 241 TLHGFLSCDLVSPESRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDALRELWQ 300
Db 241 TLHGFLSCDLVSPESRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDALRELWQ 300
Qy 301 LLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQNDKTIABG 360
Db 301 LLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQNDKTIABG 360
Qy 361 LATFFESRGWGLAKSYPADLASYYPDVLDQTANHLQAMISFSYKGGKPYMSVYLHTPEA 420
Db 361 LATFFESRGWGLAKSYPADLASYYPDVLDQTANHLQAMISFSYKGGKPYMSVYLHTPEA 420
Qy 421 FSAAAQEVAMCHDGNP 437
Db 421 FSAAAQEVAMCHDGNP 437

RESULT 2
US-09-518-657-2
; Sequence 2, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; EARLIER FILING DATE: 2000-03-03
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Neotyphodium coenophialium
US-09-518-657-2

Query Match 39.1%; Score 907.5; DB 4; Length 448;
Best Local Similarity 41.3%; Pred. No. 1.5e-86;
Matches 184; Conservative 78; Mismatches 157; Indels 27; Gaps 7;

Qy 7 AATLLPKPFVYLSQLNLNKHDTKMYSTAPMTWAGAGYDVHQAQYKFLCIHREVII 66
Db 4 AKTLHQEVYITLSETDFPANDORLWHSAPMFKQLQTANYSDAQTRHGIYKSHVI 63
Qy 67 PALGPYEPKQPMHWKSHLTRFGLPPELSFNYSKSLRFAPEPLGSLGTGKDDPNTQAI 126
Db 64 PFLGVYPTSRGE-RWLSILTRYGTPPELSLNCSDSVRYTYEPINAATGSHLDPNTFAI 122
Qy 127 RPVLQDLKAMVPGLDLEWFDHFTKALVSEEBARTLLDRDIEIPVKTQNKLAADLEPSG 186
Db 123 WEALKKHISQPGIDLEWFSYFQELTLDANESTYLHSONLVKEQIKTQNKALDLK--G 180
Qy 187 D-IVLTKYIYPRIKSIATGTPKERMFDAL--KAADKFGKVATPLAILEEFIAERAP--- 240
Db 181 DKFVLKTYIYPELKSATKSVQELVFGSVRKLAKQKHSIRPAFEMLEDYVQSRNKFTT 240
Qy 241 -----TLL-CHFLSCDLVSPESRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDA 294
Db 241 DSHNTLLSRLSLSCDLISPTKSRVKIYLLERMSVSLPAMEDLWTLGGRREDQSTIEGLEM 300
Qy 295 LRELWOLLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQND 354
Db 301 IRELWGLNNSPGLRAYPEPPLGALPNEQLPSMANYTLHNDPIPEQVYFTVFGMND 360
Qy 355 KTTAEGLATFFESRGWGLAKSYPADLASYYPDVLDQTANHLQAMISFSYKGGKPYMSVY 414
Db 414 KTTAEGLATFFESRGWGLAKSYPADLASYYPDVLDQTANHLQAMISFSYKGGKPYMSVY 414

Qy 415 LHTEF-----AFSAAAQEV 428
Db 421 LHSFETGKWPVPEGLIAFDACRRDL 446

RESULT 3
US-09-518-657-4
; Sequence 4, Application US/09518657
; Patent No. 6335188
; GENERAL INFORMATION:
; APPLICANT: Schardl, Christopher L.
; TITLE OF INVENTION: Endophyte Ergot Alkaloid Synthetic Compounds, Compounds
; FILE REFERENCE: P-1060
; CURRENT APPLICATION NUMBER: US/09/518,657
; EARLIER FILING DATE: 2000-03-03
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Neotyphodium coenophialium
US-09-518-657-4

Query Match 39.0%; Score 904.5; DB 4; Length 450;
Best Local Similarity 42.4%; Pred. No. 3.2e-86;
Matches 180; Conservative 76; Mismatches 154; Indels 15; Gaps 6;

Qy 7 AATLLPKPFVYLSQLNLNKHDTKMYSTAPMTWAGAGYDVHQAQYKFLCIHREVII 66
Db 4 AKTLHQEVYITLSETDFPANDORLWHSAPMFKQLQTANYSDAQTRHGIYKSHVI 63
Qy 67 PALGPYEPKQPMHWKSHLTRFGLPPELSFNYSKSLRFAPEPLGSLGTGKDDPNTQAI 126
Db 64 PFLGVYPTSRGE-RWLSILTRYGTPPELSLNCSDSVRYTYEPINAATGSHLDPNTFAI 122
Qy 127 RPVLQDLKAMVPGLDLEWFDHFTKALVSEEBARTLLDRDIEIPVKTQNKLAADLEPSG 186
Db 123 WEALKKHISQPGIDLEWFSYFQELTLDANESTYLHSONLVKEQIKTQNKALDLK--G 180
Qy 187 D-IVLTKYIYPRIKSIATGTPKERMFDAL--KAADKFGKVATPLAILEEFIAER--APT- 241
Db 181 DKFVLKTYIYPELKSATKSVQELVFGSVRKLAKQKHSIRPAFEMLEDYVQSRNKVPTT 240
Qy 242 -----LLGHFLSCDLVSPESRIKVCYMERQDLASTEGITWINGRNDPDTLDGLDA 294
Db 241 DSENTPLSRLSLSCDLISPTKSRVKIYLLERMSVSLPAMEDLWTLGGRREDQSTIEGLEM 300
Qy 295 LRELWOLLPVTEGLCPNCFYEPGTSPOQLPFIINFTLSPKSLPPEQIYFPAGQND 354
Db 301 IRELWGLNNSPGLRAYPEPPLGALPNEQLPSMANYTLHNDPIPEQVYFTVFGMND 360
Qy 355 KTTAEGLATFFESRGWGLAKSYPADLASYYPDVLDQTANHLQAMISFSYKGGKPYMSVY 414
Db 361 MEVTNALTKEFMREHSDMSKACLRFSFPHHNYEALNYTHSYISFSYRNKNKPYLSVY 420
Qy 415 LHTEF 419
Db 421 LHSFE 425

RESULT 4
US-09-198-452A-701
; Sequence 701, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 701
LENGTH: 430
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-701

Query Match
Best Local Similarity 4.1%; Score 94.5; DB 4; Length 430;
Matches 64; Conservative 41; Mismatches 126; Indels 51; Gaps 15;

QY 25 SNKHDKWYSAPFATMAGAG-YDVHAQYKFLCI-----HREVIIPALGYPPEKG 76
DB 104 SHRSNKKFYI-ILPFYRQAGRYQHOFQVEAIGVRHPLRDAEVLALLIDFYSRVG 162
QY 77 QPMWKSHTLRCPLPELSFNYSKLLRFAPEPLGSLGTCKDDPNTQAIR-----PVL 130
DB 163 L-CHMOIQINFLG-GSETFRDKVLRAYIKESMGELSALSOQSFSTNVRLILDSKEPED 220
QY 131 QDLKAMVPGL-----DLEWFDHFTKALVSEEEARTLLDRIDIPVFKTONKLAADLE 183
DB 221 QEIIRQAPILDVSDLDKYNELDAL-----RVLSIP-YAINPRLVRGLD 267
QY 184 PGSDIVLK-TYIYPRKSIATGTPKRLMFDKADKPKVATPLALEEPIAEAPTLL 242
DB 268 YYSDLVFEATTFQEVSYALGGGR-----YDGLISA--FGASLPACGFGVGLERAIQTL 321
QY 243 LGHFLSCDLVKFS-ESRIKVCYMERQDLASIEGIWTLLNGRR 283
DB 322 LAQ-----KRIEQFPHKRLIPNEPDADQFCLE--WSQHLRR 357

RESULT 5
US-09-489-039A-13023
Sequence 13023, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 13023
LENGTH: 630
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13023

Query Match
Best Local Similarity 4.0%; Score 93; DB 4; Length 630;
Matches 56; Conservative 25; Mismatches 64; Indels 110; Gaps 14;

QY 69 LGYPPEKGQPMHWK-----SHLTFGLPRELSFNYSKSLRFP--APEPLGSLGTGKDDPF 121
DB 320 LQRPDLPOGMEDELSERVVHVRHLPFLHATYDESISRLMDVFEKV-----NRDIEF 374
QY 122 NTOAIRPVLQDLKAMVGLDLEW-FDHFTKALVSEEEARTLLDRDIEIPVFKTONKLA 180
DB 375 N-----GLHWFFDH-----AETITERNIE-----393
QY 181 DLEPSGDIVLTKTYIPRIKSIATG-TPKRLMFDKADKPKVAT-----PLALEEPI 235
DB 394 -----RVKALGGGIAVQHRMAPQGGYFVDVRYGKEAVKHTPPVA-----KVL 434
QY 236 AERAPTLLGHFLSCDLVKPSESRIKVCYMERQDLASIEGIWTLLNGRRNDPDTLGL--- 292

DB 435 ALDVPVGLG-----TDATRVASYNPWTAL-----YWLVSGR-----TVGGWAMY 473
QY 293 -----DALRELM 299
DB 474 DDANRLPRDVALELM 488

RESULT 6
US-08-844-086-2
Sequence 2, Application US/08844086
Patent No. 5866390
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5866390el Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,086
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5030
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-844-086-2

Query Match
Best Local Similarity 4.0%; Score 92; DB 2; Length 833;
Matches 96; Conservative 65; Mismatches 192; Indels 140; Gaps 25;

QY 16 YVLSQALNSKDHDKW--WYSTAPMFATMAGAGYDVHAQYKFL-----CIHREVIIP- 67
DB 119 YDWDREVNTTDPNYKVTQWTF-----KLYEKLAYEAEVFNWVEELGTATANEVLLPD 174
QY 68 --ALGPYPEKGQPM-HWKSHLTRFG--LPPEL-SFNYSKSLRFAPEPLGSLGTGK--- 117
DB 175 GTSERGGYVVKPKRQWMLKITAYAEELLNDLDELWSESIDKMQRNWKSTGANVTF 234
QY 118 -----DDFNTQAIRPVL---QDLKAMVGLDLEWFDHFTKALVSEEEARTLLDRDIEI 169
DB 235 KVKGTDEFTVFTTRPDTLFGATFTVLAP-----EHELVDATISSEQAFAVAD--- 282
QY 170 PVFTQNKLAADLSPSGDIVLTKTYIPRIKSIATGTPKRLMFDKADKPKVATPLA 229
DB 283 --YKHQASLKSDLARTDLAKETGWYTGAYAINPNVNGKEMPINWIADYVLSYGTGAV--- 337
QY 230 ILEEFIAERAPTLLGHFLSCDLVKPSESRIKVCYMERQDLASI-----EGIWTLN 280

Db 338 -----MAVPAHQ-RDWEFAKQFDLPVEVLEGNVEEAAYTED 375
Qy 281 GRRNDPETHLGL-----DALRELWQL-----PVTEGLC-----PLPNCFY 317
Db 376 GLHVSDFDLGLNKEDAIAKIVACLEKGGQKVTYRLDWLFSRQRYWGEPIPIHWE 435
Qy 318 PGTS---PQSLPFIINFTLSPKSALPEQIYPPAFQND---KTIABGLATFFESR--- 368
Db 436 DGTSTAVPETELPLVLPVT---KDIRPSGTGESPLANLTDWLEVTREDGVKGRRETNTMP 492
Qy 369 GWGLAKSY-----PADLASYPVDVLOTAN-----HL---QAWISFSY-- 404
Db 493 QWAGSSWYLYRYIDPHNTEKLADEDLKQWLPVDIYVGAHAHLHLLYARFWHKLFLYDL 552
Qy 405 ---KGKPYMSVY 414
Db 553 GWVTEKPEPFQKLP 565

RESULT 7

US-09-018-211-2
; Sequence 2, Application US/09018211
; Patent No. 6048716
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 6048716el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,211
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 06/844,086
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: 9607993.4
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31457-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 833 amino acids
; TYPE: amino acid
; STANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-018-211-2

Query Match 4.0%; Score 92; DB 3; Length 833;
Best Local Similarity 19.5%; Pred. No. 2,2;
Matches 96; Conservative 65; Mismatches 192; Indels 140; Gaps 25;

Qy 68 ---ALGPYPEKQPM-HWKSHLTRFG--LPFEL-SFNYSKSLRFAFEPLGSLTGK--- 117
Db 175 GTSERGYPVVRKPMRQWMLKIYAYERLLNDLDELONSSIKDMQSNWIGKSTGANVTF 234
Qy 118 -----DDPNTQAIRPVL---QDLKAMVPGLDLEWFDHFTKALVVSBEERATLLDRDIEI 169
Db 235 KVGTDKEFTVFTTRPDTLFGATFTVLAP-----EHELVDAITSEQAQAVAD----- 282
Qy 170 PVFTQNKLAADLPESGDIVLKVITYPRIKSIATGTPEKRLMFDKAKADKFGKATPLA 229
Db 283 ---YKQASLSKDLARTDLAKEKTGVWTGAYAINPVNGKEMPIWIADVVLASYGTGAV--- 337
Qy 230 ILEBFIARPTLGLHFLSCDLVKPSRSRIKVCYMERQDLIASI-----EGIWTLN 280
Db 338 -----MAVPAHQ-RDWEFAKQFDLPVEVLEGNVEEAAYTED 375
Qy 281 GRRNDPETHLGL-----DALRELWQL-----PVTEGLC-----PLPNCFY 317
Db 376 GLHVSDFDLGLNKEDAIAKIVACLEKGGQKVTYRLDWLFSRQRYWGEPIPIHWE 435
Qy 318 PGTS---PQSLPFIINFTLSPKSALPEQIYPPAFQND---KTIABGLATFFESR--- 368
Db 436 DGTSTAVPETELPLVLPVT---KDIRPSGTGESPLANLTDWLEVTREDGVKGRRETNTMP 492
Qy 369 GWGLAKSY-----PADLASYPVDVLOTAN-----HL---QAWISFSY-- 404
Db 493 QWAGSSWYLYRYIDPHNTEKLADEDLKQWLPVDIYVGAHAHLHLLYARFWHKLFLYDL 552
Qy 405 ---KGKPYMSVY 414
Db 553 GWVTEKPEPFQKLP 565

RESULT 8

US-09-758-282B-155
; Sequence 155, Application US/09758282B
; Patent No. 6635463
; GENERAL INFORMATION:
; APPLICANT: Ma, Wu-Po
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamicheva, Natalie E.
; APPLICANT: Allawi, Hatim T.
; APPLICANT: Schaefer, James J.
; APPLICANT: Neri, Bruce P.
; TITLE OF INVENTION: Enzymes for the Detection of Nucleic Acid Sequences
; FILE REFERENCE: FORS 04931
; CURRENT APPLICATION NUMBER: US/09/758,282B
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 09/577,304
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 835
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-758-282B-155

Query Match 3.9%; Score 90.5; DB 4; Length 835;
Best Local Similarity 25.2%; Pred. No. 3,2;
Matches 91; Conservative 41; Mismatches 130; Indels 99; Gaps 23;

Qy 86 TRFGLPELSFNYSKSLRFP-----AFEPGLSLTGKDDPNT 123
Db 36 TSGEPVQVYGFARSLKALKEDGQAVVVPDAKAPFRHEAYEAYKAGRAPTEDFPR 95
Qy 124 Q-AIRPVLQDLKAMV-----PGLDLEWFDHFTKALVVSSE-----EARTLL-DRDIEIPVFK 173
Db 96 QLAIVKRLVDLGLVLEAPGYEA---DVLGTLAKAREGVEVAILGDRDF-----FO 148

FILING DATE: 23-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION DATA:
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/53313 (MRC-006xxx)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-422-582-8

Query Match 3.9%; Score 90; DB 4; Length 4302;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;

QY 24 LSNKDHK--WVYST---APMFATMAGAGYDVHAQYKFLCIHR-----EVIIPALG 70
DB 2487 LTKVHFECTGWDHDAEDAGAPLVALLRRCRQHCEE--FCVYKGLSSYGAVLPQFR 2544
QY 71 PYPEKGQPMHWSHLTRFGLPFELSFNYSKSLRFAPEPLGSLTGKDDPNTQAIRPVL 130
DB 2545 PHFEVGLAVVQDQGAUVVAL-----NRSALITLPEPNSATG-----LTWVL 2588
QY 131 QDLKAMV-PGL-----DLEWFDHFTKALVSEEBEARTLLDRDIEIPVKTQNK----- 177
DB 2589 HGLTASVLPGLLRQADPQHVIYSALVTLVNEYERALDVAAB--PKHERQHRAQIRKNI 2647
QY 178 -----LAADLEPSGDIVLKTYIYPRIKSIATGTPKERLMFDAIKA 217
DB 2648 ETIVSLRVHTVDDIQIQAALAOCMGPSRELVCSCCLKQTLHLK-----EAMWL--ILQ 2699
QY 218 ADKFGKVATPLAILEEFIAERAPTLG---HFLSCDLVKPSESRI-----KYC 263
DB 2700 AETTAGVTPTAIGDSIL-----NITGDLIHLASSDVRAFPQPSSELGAESPMSVQAYN 2754
QY 264 M-----EROLDIASIEGIWTLNGRNDPDTLGDALRELWQLLPVTEGL 308
DB 2755 LTSALMRILMRSEVLNEEPLTLAGEIV--AQGKSDPRSL-----LCYGG 2798
QY 309 CPLPNCFYE-----PGTSPOEQLPFI-----NFTLSPKSALEPFIYFP 348
DB 2799 APGPGCHFSIPEAFSGALANLSDVQLIELVDSNPPFGYISNYTVSTKVASMAFQTQAG 2858
QY 349 AFGQNDKTIAGLATF-----FESRGWGLAKSYPADL-----ASYYPDVLQTN--- 394
DB 2859 AQPIERLASERAITVKVPNNSDWAARGHRSANSVSVVQFQASVGAVVTLDSNPAA 2918
QY 395 --HLQAWIS-----FSYKGGKPYMSYLLHT 417
DB 2919 GLHLQNLNYLLDGHVLSERPEPVLAVYLHS 2948

RESULT 13
US-09-052-262-8
Sequence 8, Application US/09052262
Patent No. 6656681
GENERAL INFORMATION:
APPLICANT: Harris et al.
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.
STREET: One Financial Center
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.50 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/052,262
FILING DATE: Concurrently herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422,582
FILING DATE: 14-April-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9411900.5
FILING DATE: 14-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB FCT/GB94/02822
FILING DATE: 23-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326470.3
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/74118
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-052-262-8

Query Match 3.9%; Score 90; DB 4; Length 4302;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;

QY 24 LSNKDHK--WVYST---APMFATMAGAGYDVHAQYKFLCIHR-----EVIIPALG 70
DB 2487 LTKVHFECTGWDHDAEDAGAPLVALLRRCRQHCEE--FCVYKGLSSYGAVLPQFR 2544
QY 71 PYPEKGQPMHWSHLTRFGLPFELSFNYSKSLRFAPEPLGSLTGKDDPNTQAIRPVL 130
DB 2545 PHFEVGLAVVQDQGAUVVAL-----NRSALITLPEPNSATG-----LTWVL 2588
QY 131 QDLKAMV-PGL-----DLEWFDHFTKALVSEEBEARTLLDRDIEIPVKTQNK----- 177
DB 2589 HGLTASVLPGLLRQADPQHVIYSALVTLVNEYERALDVAAB--PKHERQHRAQIRKNI 2647
QY 178 -----LAADLEPSGDIVLKTYIYPRIKSIATGTPKERLMFDAIKA 217
DB 2648 ETIVSLRVHTVDDIQIQAALAOCMGPSRELVCSCCLKQTLHLK-----EAMWL--ILQ 2699
QY 218 ADKFGKVATPLAILEEFIAERAPTLG---HFLSCDLVKPSESRI-----KYC 263
DB 2700 AETTAGVTPTAIGDSIL-----NITGDLIHLASSDVRAFPQPSSELGAESPMSVQAYN 2754
QY 264 M-----EROLDIASIEGIWTLNGRNDPDTLGDALRELWQLLPVTEGL 308
DB 2755 LTSALMRILMRSEVLNEEPLTLAGEIV--AQGKSDPRSL-----LCYGG 2798
QY 309 CPLPNCFYE-----PGTSPOEQLPFI-----NFTLSPKSALEPFIYFP 348

Db 2799 APGCHFSIPEAFSGALANLSDVQLIFLVDNPPFFGVSINVTSTKVASMAFQTQAG 2858
 Qy 349 AFGQNDKTIAGLATF-----FESRGWGLAKSYPADL-----ASYYPDVLQNTAN--- 394
 Db 2859 AQPIERLASERAITVKVNNSDWAARGHRSSANSANSVVOQASVGVAVTLDSSNPAA 2918
 Qy 395 --HLQAWIS-----FSYKGGKPYMSVYLHT 417
 Db 2919 GLHLQALNYTLDDGHYLSEPEPYLAVYLHS 2948
 RESULT 14
 US-08-460-751-2
 ; Sequence 2, Application US/08460751
 ; Patent No. 5891628
 ; GENERAL INFORMATION:
 ; APPLICANT: Readers, Stephen
 ; APPLICANT: Schneider, Michael
 ; APPLICANT: Glucksmann, Sandra
 ; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
 ; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
 ; NUMBER OF SEQUENCES: 35
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/460,751
 ; FILING DATE: 02-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/413,560
 ; FILING DATE: 03-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ceruzzi, Laura A.
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7638-005
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/8741
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4303 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 US-08-460-751-2
 Query Match 3.9%; Score 90; DB 2; Length 4303;
 Best Local Similarity 20.4%; Pred. No. 51;
 Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;
 Qy 24 LSKNDHTK--WNYST----APMFATMMAGAGYDVHQAQYKFLCIHR-----EVIIPALG 70
 Db 2487 LTTKVHFECTGWHDAEAGAPLVYALLRRCRQGHCEB--FCVYKGLSSYCAVLPPGFR 2544
 Qy 71 PYPEKQPMHWKSHLTRFGLPELGFNYSKSLRFAFPLGLSLTGKDDPNTQAIRPVL 130
 Db 2545 PHFEVGLAVVVOQDLGAAVVAL-----NRSIAITLPEPNSGATG-----LTVWL 2588
 Qy 131 QDLKAMV-PGL-----DLEWDFHTKALVYVSEBEARTLLDRDIEIFVFKTNK----- 177
 Db 2589 HGI-TASVLEGLARADPOHVFYFYSIALVTNINFEVFAIDVAAF-PKHFEHRAOTIRKNIT 2647

Qy 178 -----LAADLEPSGDIVLKYIYPRIKSIATGTPKERMFDKA 217
 Db 2648 ETVLSLVHVTVDIQIQAALAQCMGFSRELVCRLKQTLHLK-----EAMML--ILQ 2699
 Qy 218 ADKFGKVATPLALEEFIAERAPTLG---HFISCDLVKPSERI-----KVYC 263
 Db 2700 AETTAGTWTPTAIGDSL-----NITGDLIHLASSDVRAPOPSLGAESPSRMVASQAVN 2754
 Qy 264 M-----ERQDLASIEGIWTNGRRNDPDTLDGLDALRELWQLLPVTEGL 308
 Db 2755 LTSALMRLMRSRVLNNEEPLTLAGEEIV--AQKRSDPRSL-----LCYGG 2798
 Qy 309 CPLENCYE-----PCTSPQEQLPFI-----NFTLSPKSALPEPQIYFP 348
 Db 2799 APGCHFSIPEAFSGALANLSDVQLIFLVDNPPFFGVSINVTSTKVASMAFQTQAG 2858
 Qy 349 AFGQNDKTIAGLATF-----FESRGWGLAKSYPADL-----ASYYPDVLQNTAN--- 394
 Db 2859 AQPIERLASERAITVKVNNSDWAARGHRSSANSANSVVOQASVGVAVTLDSSNPAA 2918
 Qy 395 --HLQAWIS-----FSYKGGKPYMSVYLHT 417
 Db 2919 GLHLQALNYTLDDGHYLSEPEPYLAVYLHS 2948
 RESULT 15
 US-09-052-469-6
 ; Sequence 6, Application US/09052469
 ; Patent No. 6380360
 ; GENERAL INFORMATION:
 ; APPLICANT: Harris et al.
 ; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner & Witcoff, Ltd.
 ; STREET: One Financial Center
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: US
 ; ZIP: 02111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk, 3.50 inch
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Wordperfect 6.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/052,469
 ; FILING DATE: Concurrently herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/422,582
 ; FILING DATE: 14-April-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9507766.5
 ; FILING DATE: 13-APR-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9411900.5
 ; FILING DATE: 14-JUN-1994
 ; PRIOR APPLICATION DATA: GB PCT/GB94/02822
 ; APPLICATION NUMBER: GB PCT/GB94/02822
 ; FILING DATE: 23-DEC-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9326470.3
 ; FILING DATE: 24-DEC-1993
 ; APPLICATION/AGENT INFORMATION:
 ; NAME: Williams, Ph.D., Kathleen M.
 ; REGISTRATION NUMBER: 34,380
 ; REFERENCE/DOCKET NUMBER: 3265/74165
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 345-9100
 ; TELEFAX: (617) 345-9111
 ; INFORMATION FOR SEQ ID NO: 6

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-052-469-6

Query Match          3.9%; Score 90; DB 4; Length 4339;
Best Local Similarity 20.4%; Pred. No. 51;
Matches 104; Conservative 67; Mismatches 175; Indels 164; Gaps 24;

QY      24 LSNKDHKK--WYST-----APMFATMMAGAGYDVHQAQYKFLCIHR-----EVIIPALG 70
Db      2444 LTKVHFECTGWHDAEDAGAPLIVALLRRRCQGHCEE--FCVYKGSLSYGAVALPPGFR 2501
QY      71 PYPEKQPMHWKSHLTRFGLPELSFNYSKSLRFAFEPLGSLTGCTKDDPNTQAIRVFL 130
Db      2502 PHFEVGLAVVQDLGAAVVAL-----NRSIAITLPEPNGSATG-----LTVWL 2545
QY      131 QDLKAMV-PGL-----DLEWFDHFTKALVVSBEARTLDRDIEIPVKTKQNK----- 177
Db      2546 HGLTASVLFGLLRQADPOHVIYSYSLAVTLVNLNEYERALDVAAE-PKHERQHEAQIRKNIT 2604
QY      178 -----LAADLEPSGDIVLKTVIYPRIKSIATGTGPKERLMFDAIKA 217
Db      2605 ETIVSLRVHTVDDIQIIAALAAQCMGPFRELVCRSCLKQTLHKL-----EAMML--ILQ 2656
QY      218 ADKFGKVATPLALEEFIAERAPTLG---HFLSCDLVKPSESRI-----KVYC 263
Db      2657 AETTAGTVPTAIGDSIL-----NITGDLIHASSDVRAPQPSSELGAESPSPRMVASQAYN 2711
QY      264 M-----EROLDIASIEGIWTLNGRRNDPETLDGLDALRELWQLLPVTEGL 308
Db      2712 LTSALMRIILMRSLVNEEBPLTAGEEIV--AQKRSDDPRSL-----LCYGG 2755
QY      309 CPLPNCFYE-----PQTSPEQLPFI-----NFTLSPKSALEPQIYVP 348
Db      2756 APGGCHFSIPEAFSCALANLSDVQLIFLVDSPNPPPGYISNYTVSTKVASMAFQTQAG 2815
QY      349 AFCQNDKTIAGLATP-----FESRGWGGLAKSYPADL-----ASYYPDVDLQTAN--- 394
Db      2816 AQIPIERLASERAITVKVPNNSDWAARGHRSSANSANSVVQFQASVGAVVTLDSNPAA 2875
QY      395 --HLQAWIS-----FSYKGGKKPYMSVYLHT 417
Db      2876 GLHLQLNYTLTDGHYLSSEPEPYLAVYLHS 2905
```

Search completed: July 22, 2004, 18:46:58
Job time : 21 secs

Sequence 19, Appl
Sequence 3347, Ap
Sequence 168, App
Sequence 72824, A
Sequence 66046, A
Sequence 43, Appl
Sequence 36, Appl
Sequence 121262,
Sequence 3386, Ap
Sequence 149090,
Sequence 54706, A
Sequence 701, App
Sequence 23, Appl
Sequence 30, Appl
Sequence 10341, A
Sequence 241638,
Sequence 38939, A
Sequence 159892,
Sequence 3, Appli
Sequence 74920, A
Sequence 11, Appl
Sequence 20151, A
Sequence 7965, Ap
Sequence 7966, Ap
Sequence 308, App
Sequence 744, App
Sequence 2474, Ap
Sequence 13226, A
Sequence 795, App
Sequence 65022, A

16 102 4.4 1016 9 US-09-986-224-19
17 101.5 4.4 930 15 US-10-369-493-3347
18 99.5 4.3 3234 15 US-10-093-463-168
19 98.5 4.2 507 12 US-10-425-114-72824
20 98.5 4.2 769 12 US-10-282-122A-66046
21 98 4.2 1939 12 US-10-152-886-43
22 97.5 4.2 344 12 US-10-617-038-36
23 96.5 4.2 1095 16 US-10-437-963-121262
24 95.5 4.1 503 15 US-10-369-493-3386
25 95 4.1 876 16 US-10-437-963-149090
26 94.5 4.1 430 12 US-10-282-122A-54706
27 94.5 4.1 430 15 US-10-289-762-701
28 94.5 4.1 453 16 US-10-448-871A-23
29 94.5 4.1 578 8 US-08-808-031A-30
30 94 4.0 466 14 US-10-156-761-10341
31 93.5 4.0 818 12 US-10-424-599-241638
32 93.5 4.0 818 12 US-10-425-114-38939
33 93 4.0 714 16 US-10-437-963-159892
34 93 4.0 2492 9 US-09-991-358-3
35 92.5 4.0 585 12 US-10-282-122A-74920
36 92.5 4.0 862 14 US-10-267-989-11
37 92.5 4.0 3250 15 US-10-369-493-20151
38 92 4.0 443 12 US-10-335-977-7965
39 92 4.0 448 12 US-10-335-977-7966
40 92 4.0 491 10 US-09-934-455-308
41 92 4.0 491 12 US-10-225-066A-744
42 92 4.0 491 15 US-10-374-780A-2474
43 92 4.0 597 9 US-09-815-242-13226
44 92 4.0 608 12 US-10-087-192-795
45 91.5 3.9 570 12 US-10-282-122A-65022

ALIGNMENTS

RESULT 1
US-10-099-704-2
; Sequence 2, Application US/10099704
; Publication No. US20020197682A1
; GENERAL INFORMATION:
; APPLICANT: Christensen, Bjorn Eggert
; APPLICANT: Mollgaard, Henrik
; APPLICANT: Kaasgaard, Svend
; APPLICANT: Lembeck, Jan
; TITLE OF INVENTION: Methods for producing polypeptides in
; FILE OF INVENTION: aspergillus mutant cells
; FILE REFERENCE: 4483.200-US
; CURRENT APPLICATION NUMBER: US/10/099,704
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/472,364
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/139,593
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/117,396
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: PA 1999 00745
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: PA 1998 01726
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer 5956
US-10-099-704-2

Query Match 100.0%; Score 2322; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.5e-224;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:46:15 ; Search time 45 Seconds
(without alignments)
3041.075 Million cell updates/sec

Title: US-10-099-704-2

Perfect score: 2322

Sequence: 1 MEISKAAATLLPKPFVLSQ.....FEAFSAAAEVAMCHDGHNP 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2322	100.0	437	13	US-10-099-704-2
2	402	17.3	330	15	US-10-369-493-12970
3	346.5	14.9	430	15	US-10-369-493-13332
4	313.5	13.5	249	15	US-10-369-493-12523
5	293.5	12.6	379	15	US-10-369-493-3245
6	114	4.9	577	16	US-10-437-963-182042
7	110	4.7	477	12	US-10-425-114-69165
8	104.5	4.5	393	16	US-10-437-963-122089
9	104	4.5	524	15	US-10-369-493-2592
10	103.5	4.5	880	14	US-10-059-909-12
11	102	4.4	521	9	US-09-848-035-6
12	102	4.4	521	9	US-09-986-224-6
13	102	4.4	605	15	US-10-407-866-74
14	102	4.4	635	15	US-10-161-927-4
15	102	4.4	662	15	US-10-407-866-101

QY 1 MEISKKAATLLPKPFVYVLSQALNSKNDHTKMWYSTAPMFATMMAGAGYDVHAQYKFLCI 60
DB 1 MEISKKAATLLPKPFVYVLSQALNSKNDHTKMWYSTAPMFATMMAGAGYDVHAQYKFLCI 60
QY 61 HREVIIPALGPYEPKGMWKSHTLRFGLPFELSFNYKSLRFAFEPLGSLTGTCKDP 120
DB 61 HREVIIPALGPYEPKGMWKSHTLRFGLPFELSFNYKSLRFAFEPLGSLTGTCKDP 120
QY 121 FNTQAIRPVLDLQKAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVKTQNKLA 180
DB 121 FNTQAIRPVLDLQKAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVKTQNKLA 180
QY 181 DLPSGDVILKTYIYPRISATGTPKERLMFDAIKAADKFGKVATPLAILEEFIAERAP 240
DB 181 DLPSGDVILKTYIYPRISATGTPKERLMFDAIKAADKFGKVATPLAILEEFIAERAP 240
QY 241 TLGHFLSCDLVPSRSRIKVMYCMERQDLASIEGWTINGRNDPDTLDGLDALRELWQ 300
DB 241 TLGHFLSCDLVPSRSRIKVMYCMERQDLASIEGWTINGRNDPDTLDGLDALRELWQ 300
QY 301 LLPVTEGLCLPNCFCYEPGTSPOEQLPFIINFTLSPKSLPPEQIYFPAFGQNDKTIABG 360
DB 301 LLPVTEGLCLPNCFCYEPGTSPOEQLPFIINFTLSPKSLPPEQIYFPAFGQNDKTIABG 360
QY 361 LATPFSSRGWGLAKYPADLASYPDVLQTNHLQAWISFSYKSKKPYNSVYLHTFEA 420
DB 361 LATPFSSRGWGLAKYPADLASYPDVLQTNHLQAWISFSYKSKKPYNSVYLHTFEA 420
QY 421 FSAAGAEVAMCHDGNP 437
DB 421 FSAAGAEVAMCHDGNP 437

RESULT 2
US-10-369-493-12970
; Sequence 12970, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12970
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Aspergillus nidulans

Query Match 17.3%; Score 402; DB 15; Length 330;
Best Local Similarity 31.3%; Pred. No. 2e-31;
Matches 105; Conservative 61; Mismatches 125; Indels 44; Gaps 12;

QY 10 LIPK-----PFYVLSQALNSKNDHTKMWYSTAPMFATMMAGAGYDVHAQYKFLCIHREY 64
DB 1 LVPKPISTQPTLLTSLYFTNEAQWLDGSLFSRLQASQYTTGQYKHLFFHRY 60
QY 65 IIPALGPYEPKGMWKSHTLRFGLPFELSFNYKSLRFAFEPLGSLTGTCKD--D 119
DB 61 LIPFLGTYPT-----WFSIVSRQGLAIEYLNFOAGNPFVRSAPFELSVASG-KDAAD 114
QY 120 PNTQAIRPVLDLQKAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVKTQNK 178
DB 115 PLNKRTEDLLAGLVEQGLEGDMTLFDHVRTITISDEETALENPDAY--GKHTAL 172

QY 179 AADLEPSGD-IVLKTIVYPRISATGTPKERLMFDAI-KAADKFGKVATPLAILEEFIA 236
DB 173 GFDWK--GDNVTVKCYNHPKWSLATGVPVAKLRDSLERIKQOF-DCEQALVELVDEIME 229
QY 237 ERAPTLGHFLSCDLVPSRSRIKVMYCMERQDLASIEGWTINGRNDPDTLDGLDALR 296
DB 230 DSGSWDLRTFIAMDVCPLAQTELKIYGIINEVSLGKVEELWTMGRLNDETTLGLESLR 289
QY 297 ELWOLL-----PVTGLCPL 311
DB 290 RLHLLEVNKDDRLFSKGBEKLEYGPTVDGFLPL 324

RESULT 3
US-10-369-493-13332
; Sequence 13332, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13332
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Aspergillus nidulans

Query Match 14.9%; Score 346.5; DB 15; Length 430;
Best Local Similarity 26.3%; Pred. No. 1.1e-25;
Matches 120; Conservative 55; Mismatches 177; Indels 105; Gaps 15;

QY 17 VLSQALNSKNDHTKMWYSTAPMFATMMAGAGYDVHAQYKFLCIH----- 61
DB 7 VLSQTLPSRGPDVDAMWQLTGRHLAVLLDAAAYPIEKQYECCLLYHYHAYVCLTTSYKVT 66
QY 62 ---REVLIIPALGPYEPKGMWKSHTLRFGLPFELSFNYKSLRFAFEPLGSL 113
DB 67 DXLEVXAPYLGAPAPREGASPTWKSMQLDGPFFEFSSKWNPNPGGPDVRFGLSEIGPM 126
QY 114 TGTKDDPNTQAIRPVLDLQKAMVPGDLLEWDFHFTKALVVSSEEARLTDRIEIPVK 173
DB 127 AGTSLDPLNHLAMKEILYKLSAVPGSGLTWTTHF-----LATLFDH-----YA 171
QY 174 TQNKLAADLEPS-----GDVILKTYIYPR-----IKSIATG 204
DB 172 KYTQKAATMGSSITGTSIWSLEFQRKSTGLKTYFHPKLDQQQAFLDIPSWAESFRGLHEN 231
QY 205 TPKERLMFDAIKAADKFGKVATPLAILEEFIAERAPTLGHFLSCDLVPSRSRIKVMYCM 264
DB 232 SPRTAVHEFL-STNPEGLKKPFC-----LSVNCSPAKARIKMYEN 273
QY 265 BROLDLASIEGWTINGRNDPDTLDGLDALRELWOLL-PVTBGLCLPNCFCYEPGTSQ 323
DB 274 SPHTNFRAIREINTLGGRIADTET--RTKQFSLEFLNLLKTVTRGARXFFPRDF-RVSIQ 330
QY 324 EQ-----LPIFIN-----FTLSPKSLPPEQIYFPAFG--QNDKTIAGELATFFES 367
DB 331 QRQHNPQLDADPMLKGCYVFFDIAPGNLPAIKYFFPVNRHCRNDLAVTQNLNEWLES 390


```
QY 101 SLRFAPEPLGS-LTGTKDDP--FNTQAIRPVLQ-DLKAMVP-----GLDLEWFDHF 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 TWVR-AKHPLHSALTGWCECFDFPLQA--PGLQAGLDTVPSSKVLKNSLDLTRYDY 281
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 TKALVSEEBARTLLDRDIEIPVFKTONKLAADLEPSGD-----IVLKIYIP- 196
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 T-----VGINKYTERYVDNPAGDYHRLRYDYGNYFASKTFDFDV 322
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 RIKSIATGTGKERLMEFDAIKADKFGKVATPLAI-----LEEPIAERAPTLIGHFLS 248
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 KERRILLGWANESDVTYDKAKGWAGIHALPRKVLDPDGKQLQWPPIBELSLTRKSVS 382
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 249 C--DLVKPSE--SRIKVY--CMERQDLASIEGIWTLN-GRNDPDTLDG----- 292
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 VFDKWKPGSHFQVGTGLTYQADVEVSLVSGLEKAEALDPAFGDDAERLCGAKGADV 442
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 DALREIWLPLVTEGLCPLENCF---VEPGTSPQEQLPFLINFTLSPKSPALPEPOLY 349
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 GVVFGLVWL--ASAGLEEXTAVFRVFKPACHGAK--PVVLCTDPTKSSL-SPDLY 497
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 350 F-GONDKTIAEGLATF-----FESRGWG--LAKSYPA-----DLASY--PDV 388
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 FAGFVDTDISSGKISLRSILDRWSBESFGAGGKTCILSRVYFSMAIGDKAHLVFN 557
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 389 DLOTANHLOAWISFSYKGGKPYNS 412
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 558 DIK-ISHLKAW-----EMKKPLMN 575
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 7

```
US-10-425-114-69165
; Sequence 69165, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69165
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMELMO17129G02_F11.pep
US-10-425-114-69165
```

```
Query Match 4.7%; Score 110; DB 12; Length 477;
Best Local Similarity 26.4%; Pred. No. 0.086;
Matches 55; Conservative 34; Mismatches 81; Indels 38; Gaps 10;

QY 118 DDFFNTQAIRPV-LQDLKAMVGLDLEWFDHFTKALVSEEBARTLLDRDIEIPV----- 171
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 197 DDDATTPKARKVKLVKNVWAIAGLAK-----TTKALVETSE-----LVTNSSPVTVDP 248
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 172 -----FKTONKLAADLEPSGDIVL---KTYIYPRIKSIATGTPXERLMFMDAIAADKFG 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 HGTLHRAKTQDNQNSAFGGEKFMIRGKTYLTLDYHKVVGDPPLLELLAVDFKNERFD 308
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 KVAT-PLAILEEPIAERAPTLIGHFLSCDLVKPSSRIKVCYMERQDLASIEGIWTLNG 281
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 SVAHFKSLVQSEAARKGPFILVINLQVP-AKENTNLVWYAAERFVNKDSLLG----- 361
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 RRNDPETHGLDALREL-WQLLP-VTEG 307
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 -----RFIDGTDAYRDAREFKLIPSIIVEG 384
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 8

```
US-10-437-963-122089
; Sequence 122089, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122089
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_2504C.1.pep
US-10-437-963-122089
```

```
Query Match 4.5%; Score 104.5; DB 16; Length 393;
Best Local Similarity 21.6%; Pred. No. 0.23;
Matches 106; Conservative 59; Mismatches 160; Indels 165; Gaps 27;

QY 7 AATLLP---KPFYVLSQALNLSNKHDTKWYSTAPMPATPMAGAGYDVHAYKFLCIHRE 63
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 AARLLPVAARLTTTARARLSTST-----STSPATAVL-----YDQH----- 42
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 VIIPALGPYPEKGQPMHWKSHLTFGLPPELSFNYSKSLRPAFEPGLSLTGTKD----- 118
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 43 -----GP-PDK-----VLRAELPAAEI-GERDVCVRM 68
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 --DPFN-----TQAIRPVLQDLKAMVGLDLEWFDHFTKALVSEEBARTLLDRDIEIP 170
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 LAAPINPDLNRVEGVYFVRPPLPAAVAGYGVGVHALGGAV-----DSRLLSPGDWVIP 124
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 171 VFRTONKLAADLBSPGDIVLTKTYI-----YPRKS-----TATGTPKE-----RLMFEA 214
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 -----SPPSLGTWTQYIVAPATAWHVRSDVPPQYVATVTNPLTALRMLCDF 172
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 215 IKAADKFGKVATPLAILEEPI---ABRAPTLGLGHFLSCD-LVKPSSRIKVCYMERQDL 270
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 VNLAPGL-----LSLLSLFFPCNQVSPALDTFVSGDTLVQNGATSIVGQCV---IOL 223
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 ASIEGIWTLNGRNDPETHGLDALRELWQLLPVTEGLCPNCFYEPGTSPOEQLPFI 330
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 AKHLGLHTINIIRDRPGSQBAKDKLQLGADHVFTESQLDIRN----- 266
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 NFTLSPKSALPEPIQYFPFAPGQNDKTIAGLAFVFFESRG-----WGLAKSYPADLAS--- 383
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 --IKSLIGALPEPALGLNLCVGGNAASV---ILKFLRQGGTMTVYGGMSKK-PVTVTSS 320
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 384 YYPDVLQTNANHLOANIS-----FSYKHKPYMSVYLHTTFEAFSAAAE 427
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 IFKDLSLR-GFWLQKWMSSDKAESRTMIDYLLDLVHEGLKLY-EMELTFFSDFHLDLK 378
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 428 VAMCHDGHNP 437
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 -ALGKGSQP 387
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 9

```
US-10-369-493-2592
; Sequence 2592, Application US/10369493
```

```

; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 880
; TYPE: PRT
; ORGANISM: Momordica charantia
US-10-059-909-12

Query Match
Best Local Similarity 4.5%; Score 103.5; DB 14; Length 880;
Matches 96; Conservative 46; Mismatches 138; Indels 135; Gaps 24;

QY 49 YDVHAQYKFLCIHR-----EVIIPALG-----PYPEK---GQPMHWKSHLTRFGLPFELSEN 97
DB 221 YDV---YNDLCDPNGGPNLVRPILGSDQYPPRRGRTRGRPPARKDKHYESRLSDVMSLN 277
QY 98 -YSKLLRAFAPEPLGSLGTGDDPNTQAIRVLQDLKAMVPGDLEWPDHDTKALVUSE 156
DB 278 IYVRDENFGHLKMADFLGNTLKLVLST-SIQGLSEIFDSTFG---EFDKF----- 324
QY 157 BEARTLLDRDIBP--VEKTONKLAADLEPSGDIVLKYIYPRIKSIATGTGTERLMFDA 214
DB 325 KEVDOLFGRGPPILNIFK---NLTTEDLAPP-----LFKA 356
QY 215 IKAAD--KFGKVATLAILE-----EFIAERAPTLGHFLSCDLVKPSESRI 259
DB 357 FLRSDGRFLKYPTQVIKDNKLGWRTDEEFAREMIAGVNPILII-----RRL 403
QY 260 KYVCHERQLDLASIEGIWTNGRRNDPET-----LDGL---DALRE-LWQLLPVTEGL 308
DB 404 EVFFPLSKLDP-----HYVGNQNSTWTEEQIKHGLDGLTVDEAIKENKLYILDHHDAL 456
QY 309 CPLNCFEYPGTSPQEQLPFIIT---NFTLSP---KSALPEPQ-----IYFFAFGQN 353
DB 457 MEYLARINSTKTYATRTLLFLKDDSTLKPLAIELSLPHPOGDHGAISKLYFFAEGR- 515
QY 354 DKTIAGLANTPESRGWGGLAKSPADLASIYPPDVS---LQTANHLOAWISFSYK 405
DB 516 -----VESAIW-QLAKAYVAVNDSGYHQLNSHWLTHAVLEPEVITTHR 558

RESULT 11
US-09-848-035-6
; Sequence 6, Application US/09848035
; Patent No. US20020028465A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE NBS/LRR PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-268001
; CURRENT APPLICATION NUMBER: US/09/848,035
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/201,464
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-848-035-6

Query Match
Best Local Similarity 4.4%; Score 102; DB 9; Length 521;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;

QY 79 MHWKSHL---TRFGLPFELSFNYSKSLRFAFEPGLSGTGT-K-DDPFNTQAIRPVLQDJK 134
DB 67 LHWANGVLFPQRFYSYVFLSCHIKRYMKETTFAELISLDWDPDFAPIEBFMSQP--EKLK 124
QY 135 AMVPGLDLEWDFHFTKALVUSSEEARTLDRD-----IEIPVFK-TQNKLAADLEPSGD 187
DB 125 FIIDGFE-----EIIISERSGESLDDGSPCTDWQOELSVTKILHSLKKELPLAT 175

```

QY 188 --IVLKYIYPRKSI-----ATGTPKERLMFDAIKAAKFGKVATPLAILEE----- 233
Db 176 LLITIKTWFRDLKASLVNFCFVQITGTTGDDLRVVFMRHFDSDSSEVEKILQQLRKNETL 235
QY 234 FIABRAPTLGHLFSLC-----DLVKPSESRIKYVC-----MERQLDLA----- 271
Db 236 FHSCSAPMVCWTVCSCLKQKRVYDLSQITQTTSLYAYFFSNLSTAEVDLADDSWPG 295
QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 296 QWRALCSLAIEGLWSMNFNFKEDTEIEGLEVPFIDSLYEFNLQKINDCGGCTTFTHLS 355
QY 312 -----PNCIFYEPGTSPOQLPFIINFILSPKLSALPEPOLYFPFAGQNDKTI 357
Db 356 FOEFAAMS FVLEBEPFPFHSTKPOE-MKQLLQHVLLDKEAYWTPVVLFP-FGLLNKNI 413
QY 358 AEGL 361
Db 414 AREL 417

RESULT 12
US-09-986-224-6
; Sequence 6, Application US/09986224
; Patent No. US20020150920A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; APPLICANT: Wang, Weiye
; TITLE OF INVENTION: NOVEL MOLECULES OF THE NBS/LRR PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-333001
; CURRENT APPLICATION NUMBER: US/09/986,224
; 2001-10-22
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/848,035
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/201,464
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-224-6

Query Match 4.4%; Score 102; DB 9; Length 521;
Best Local Similarity 22.3%; Pred. No. 0.63;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPELSEFNYSKSLRFAPEPLGSLTGK-DDPFTQAIRPVLQDLK 134
Db 67 LHWANGVLFOQRFYSYVFLSCHIRYKETTFAELISLDWPDFDAPIEFMSQP--EKLL 124
QY 135 ANVPGLDLEWFDHFTKALVVSBEARTLDRD-----LEIPVK-TQNKLAADLEPSGD 187
Db 125 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTIKILHSLKKEVLPLAT 175
QY 188 --IVLKYIYPRKSI-----ATGTPKERLMFDAIKAAKFGKVATPLAILEE----- 233
Db 176 LLITIKTWFRDLKASLVNFCFVQITGTTGDDLRVVFMRHFDSDSSEVEKILQQLRKNETL 235
QY 234 FIABRAPTLGHLFSLC-----DLVKPSESRIKYVC-----MERQLDLA----- 271
Db 236 FHSCSAPMVCWTVCSCLKQKRVYDLSQITQTTSLYAYFFSNLSTAEVDLADDSWPG 295
QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 296 QWRALCSLAIEGLWSMNFNFKEDTEIEGLEVPFIDSLYEFNLQKINDCGGCTTFTHLS 355
QY 312 -----PNCIFYEPGTSPOQLPFIINFILSPKLSALPEPOLYFPFAGQNDKTI 357

QY 358 AEGL 361
Db 414 AREL 417

RESULT 13
US-10-407-866-74
; Sequence 74, Application US/10407866
; Publication No. US20040002593A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides, and Methods of Use
; FILE REFERENCE: 66654-10(LJ 5755)
; CURRENT APPLICATION NUMBER: US/10/407,866
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: US 60/370,538
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-407-866-74

Query Match 4.4%; Score 102; DB 15; Length 605;
Best Local Similarity 22.3%; Pred. No. 0.79;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
QY 79 MHWKSHL---TRFGLPELSEFNYSKSLRFAPEPLGSLTGK-DDPFTQAIRPVLQDLK 134
Db 232 LHWANGVLFOQRFYSYVFLSCHIRYKETTFAELISLDWPDFDAPIEFMSQP--EKLL 289
QY 135 ANVPGLDLEWFDHFTKALVVSBEARTLDRD-----LEIPVK-TQNKLAADLEPSGD 187
Db 290 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTIKILHSLKKEVLPLAT 340
QY 188 --IVLKYIYPRKSI-----ATGTPKERLMFDAIKAAKFGKVATPLAILEE----- 233
Db 341 LLITIKTWFRDLKASLVNFCFVQITGTTGDDLRVVFMRHFDSDSSEVEKILQQLRKNETL 400
QY 234 FIABRAPTLGHLFSLC-----DLVKPSESRIKYVC-----MERQLDLA----- 271
Db 401 FHSCSAPMVCWTVCSCLKQKRVYDLSQITQTTSLYAYFFSNLSTAEVDLADDSWPG 460
QY 272 -----STEGIWTNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 461 QWRALCSLAIEGLWSMNFNFKEDTEIEGLEVPFIDSLYEFNLQKINDCGGCTTFTHLS 520
QY 312 -----PNCIFYEPGTSPOQLPFIINFILSPKLSALPEPOLYFPFAGQNDKTI 357
Db 521 FOEFAAMS FVLEBEPFPFHSTKPOE-MKQLLQHVLLDKEAYWTPVVLFP-FGLLNKNI 578
QY 358 AEGL 361
Db 579 AREL 582

RESULT 14
US-10-161-927-4
; Sequence 4, Application US/10161927
; Publication No. US20030235821A1
; GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Miller, Charles E.
; APPLICANT: Hjalte, Tord

APPLICANT: Baumgartner, Jason C.
APPLICANT: Guo, Xiaojia
APPLICANT: Gangolli, Esha A.
APPLICANT: Vernet, Corine
APPLICANT: Padigar, Muralidhara
APPLICANT: Li, Li
APPLICANT: Pena, Carol E.A.
APPLICANT: Gorman, Linda
APPLICANT: Anderson, David W.
APPLICANT: Edinger, Shlomit R.
APPLICANT: Patturajan, Meera
APPLICANT: Stone, David J.
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
FILE REFERENCE: 21402-377 D (Cura 677 Other)
CURRENT APPLICATION NUMBER: US/10/161,927
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: 60/295,661
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/295,607
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/296,404
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,418
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,575
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: 60/297,414
PRIOR FILING DATE: 2001-06-11
PRIOR APPLICATION NUMBER: 60/297,567
PRIOR FILING DATE: 2001-06-12
PRIOR APPLICATION NUMBER: 60/298,528
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/325,685
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/299,133
PRIOR FILING DATE: 2001-06-18
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 190
SEQ ID NO 4
LENGTH: 635
TYPE: PRT
ORGANISM: Homo sapiens
US-10-161-927-4

Query Match 4.4%; Score 102; DB 15; Length 635;
Best Local Similarity 22.3%; Pred. No. 0.85;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
Qy 79 MHKSHL---TRGLPELSFNYSKSLRFAPEPLGSLGTK-DDPNTQAIRPVLODLK 134
Db 67 LHWANGVLFQORFSYFVLSCHKIRYMKETFAELISLDWPDFDAPIEFMSQP--EKLL 124
Qy 135 AMVPGLDLEWDFHFTKALVSEBEARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
Db 125 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTKILHLLKELVPLAT 175
Qy 188 --IVLKYIYPRIKSI-----ATGTPKERMFDAKAADKFGKVATPLAILEE---- 233
Db 176 LLITIKTWFRDLKASLVNCFVQITGTDGLRVYFMHRHFDSDSEVEKILQOLRKNETL 235
Qy 234 FIAERAPTLGHFLSC-----DLVKPSESRIKYC-----MERQLDLA----- 271
Db 236 FHSCAPMVCWVCSCLKQPKVRYDIQSITQTTSIYAFFSNLFTAEDVLADDSWPG 295
Qy 272 -----SIEGIWTLNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 296 QWRALCSLAIEGLWMNFTNKEDTEIEGLEVPFIDSLYEFNLIQKINDCGGCTTFTHLS 355
Qy 312 -----PNCFYEGTSPQOLPFIINFTLSPKSALPEPQIYFPAPGQNDXTI 357
Db 356 FQEFFAAMSVLEBPREFPHSTKPOB-MKMLLQHVLLDKKEYWTPVVLFP-FFGLLNKNI 413

Qy 358 AEGL 361
Db 414 AREL 417
RESULT 15
US-10-407-866-101
; Sequence 101, Application US/10407866
; Publication No. US20040002593A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; FILE REFERENCE: 66654-10(LJ 5755)
; CURRENT APPLICATION NUMBER: US/10/407,866
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: US 60/370,538
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-407-866-101
Query Match 4.4%; Score 102; DB 15; Length 662;
Best Local Similarity 22.3%; Pred. No. 0.9;
Matches 81; Conservative 51; Mismatches 138; Indels 94; Gaps 18;
Qy 79 MHKSHL---TRGLPELSFNYSKSLRFAPEPLGSLGTK-DDPNTQAIRPVLODLK 134
Db 232 LHWANGVLFQORFSYFVLSCHKIRYMKETFAELISLDWPDFDAPIEFMSQP--EKLL 289
Qy 135 AMVPGLDLEWDFHFTKALVSEBEARTLLDRD-----IEIPVK-TQNKLAADLEPSGD 187
Db 290 FIIDGFE-----EIIISERSSELDGSPCTDWYQELPVTKILHLLKELVPLAT 340
Qy 188 --IVLKYIYPRIKSI-----ATGTPKERMFDAKAADKFGKVATPLAILEE---- 233
Db 341 LLITIKTWFRDLKASLVNCFVQITGTDGLRVYFMHRHFDSDSEVEKILQOLRKNETL 400
Qy 234 FIAERAPTLGHFLSC-----DLVKPSESRIKYC-----MERQLDLA----- 271
Db 401 FHSCAPMVCWVCSCLKQPKVRYDIQSITQTTSIYAFFSNLFTAEDVLADDSWPG 460
Qy 272 -----SIEGIWTLNGRNDPET-LDGL-----DALRELWQLLPVTE-GLCPL----- 311
Db 461 QWRALCSLAIEGLWMNFTNKEDTEIEGLEVPFIDSLYEFNLIQKINDCGGCTTFTHLS 520
Qy 312 -----PNCFYEGTSPQOLPFIINFTLSPKSALPEPQIYFPAPGQNDXTI 357
Db 521 FQEFFAAMSVLEBPREFPHSTKPOB-MKMLLQHVLLDKKEYWTPVVLFP-FFGLLNKNI 578
Qy 358 AEGL 361
Db 579 AREL 582
Search completed: July 22, 2004, 18:51:49
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:41:47 ; Search time 17 Seconds
(without alignments)
2472.688 Million cell updates/sec

Title: US-10-099-704-2

Perfect score: 2322
Sequence: 1 MEISKATLLPKPFYLSQ.....FEAFSAAQEVAMCHDGNP 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	895.5	38.6	455	1 JC4338	tryptophan dimethyl
2	111.5	4.8	378	2 C64091	adenine glycosylas
3	111.5	4.8	555	1 SYHQMA	malate synthase [E
4	107.5	4.6	490	2 T23791	hypothetical prote
5	101	4.3	878	2 S74207	lipoxigenase [EC 1
6	101	4.3	1621	2 A82255	hypothetical prote
7	100	4.3	323	2 G87379	heptosyltransferas
8	99.5	4.3	486	2 JC4240	lanosterol 14alpha
9	99.5	4.3	503	2 JC4758	transcription init
10	98.5	4.2	503	2 G85432	GTP pyrophosphokin
11	98.5	4.2	769	2 D81828	hypothetical prote
12	97.5	4.2	344	2 F70322	aldehyde-ferradoxi
13	96.5	4.2	621	2 A71221	hypothetical prote
14	95.5	4.1	1073	2 S69079	genome polyprotein
15	95.5	4.1	1882	1 GNVVTR	histidyl tRNA synt
16	95	4.1	430	2 D86573	oligopeptide ABC t
17	95	4.1	486	2 B83009	histidine-tRNA lig
18	94.5	4.1	430	2 H72052	pol polyprotein -
19	94.5	4.1	896	1 GNLJQH	hypothetical prote
20	94	4.0	910	2 S67034	hypothetical prote
21	93.5	4.0	693	2 T33251	leucyl-tRNA synth
22	93	4.0	833	2 H95029	nonstructural poly
23	93	4.0	2492	1 MNWVTD	dihydroterocate sy
24	92.5	4.0	476	2 E69426	DNA-packaging prot
25	92.5	4.0	640	2 T13088	aldehyde dehydroge
26	92.5	4.0	862	1 A49346	ES4F2.8 protein -
27	92	4.0	282	2 S44825	hypothetical prote
28	92	4.0	374	2 AD2802	periplasmic bindin
29	92	4.0	379	2 E97581	

30 92 4.0 443 2 F71929 heat shock protein
31 92 4.0 534 2 T04663 hypothetical prote
32 92 4.0 1017 2 S62435 probable glycine d
33 92 4.0 2492 1 C44213 nonstructural poly
34 91.5 3.9 466 2 T40150 probable glycylnep
35 91.5 3.9 585 2 AG0955 conserved hypotet
36 91.5 3.9 986 2 T41809 DNA polymerase orf
37 91.5 3.9 987 2 AF2296 hypothetical prote
38 91 3.9 399 2 S75523 mannosyltransferas
39 91 3.9 470 2 S50083 photolyase - short
40 91 3.9 775 2 S28284 hypothetical prote
41 91 3.9 833 2 C97901 leucine-tRNA ligas
42 90.5 3.9 521 2 D64434 serine-tRNA ligase
43 90.5 3.9 586 2 JC2407 homothallic switch
44 90.5 3.9 852 2 S41886 DNA repair protein
45 90.5 3.9 1736 2 T05174 hypothetical prote

ALIGNMENTS

RESULT 1

JC4338
tryptophan dimethylallyltransferase (EC 2.5.1.34) - ergot fungus
N/Alternate names: dimethylallyl diphosphate:l-tryptophan dimethylallyltransferase; dim
C/Species: Claviceps purpurea (ergot fungus)
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: JC4338
R/Tsai, H.F.; Wang, H.; Gebler, J.C.; Poulter, C.D.; Schardl, C.L.
Biochem. Biophys. Res. Commun. 216, 119-125, 1995
A/Title: The Claviceps purpurea gene encoding dimethylallyltryptophan synthase, the comm
A/Reference number: JC4338; MUID:96067540; PMID:7488077
A/Accession: JC4338
A/Molecule type: mRNA
A/Residues: 1-455 <TSA>
A/Cross-references: GB:L39640; NID:g1005417; PID:g1005418
A/Experimental source: ATCC 26245
C/Genetics:
A/Gene: dmaW
A/Introns: 396/2; 437/1
C/Function:
A/Description: catalyzes the synthesis of 4'-(gamma,gamma-dimethylallyl)-tryptophan fro
A/Pathway: ergot alkaloid biosynthesis
C/Superfamily: ergot fungus tryptophan dimethylallyltransferase
C/Keywords: alkaloid biosynthesis; transferase
F/113-117/Region: prenyl diphosphate binding #status predicted

Query Match 38.6%; Score 895.5; DB 1; Length 455;
Best Local Similarity 42.1%; Pred. No. 2.2e-64;
Matches 187; Conservative 75; Mismatches 151; Indels 31; Gaps 9;

QY 3 ISKKAATLLPKPFYVLSQALNSKDHDKWVYSTAPMEATMWAGAGYDVHAQYKFLCIHR 62
DB 2 MTKAPATAV---YDTLSLLDFPNQEQRLMWHSTAPMAALDDTAGHNVHDOYRELGLFK 58
QY 63 EVIIPALGPYEPKQPMH-WKSHLTRFGLPELSFNYSKSLRFAFEPLGLSTGKDDPF 121
DB 59 KHIIFPLGYPAQK--HTWPSVLTRYGIPELSLNCILDSVVRYTFEPTTEHTGTGDDSY 116
QY 122 NTAQTRPVLQDKAMVPGLDLEWPHFTKALVWSEEARLTILDRD-IEIPVKTOKLAA 180
DB 117 NAFALTECIQKLVRIQPGIDMEWFSYFNNELVNLNATESARLGRNDSVNVQPIRTOKLAL 176
QY 181 DLEPSGD-IVLKYTYIPRIKSTATGTPKRLMFDAL-KAADKFGKVATPLAILEEFFIABR 238
DB 177 DUK--GDRFALKVLYLPHLKSATGVSSHDLIFNSVRKLSQKHTSIQSFNVLCDYVASR 234
QY 239 -----APTLLGHFLSCDLVVKSESRKIVYCHMERQDLIASIGIWTINGR 283
DB 235 NPDPSNAAEAGVPASRALRLLSCDLVDPSKSIKTYLLEQTVSLTAMEDLWLGRR 294
QY 284 NDEPTLDGLDALRELWQLLPVTEGLCPLENCPEYEGTSPQEQLPFIINFTLSPKSALPEP 343

Db 295 TDSSTLNGLDMMRELHLLQIPSGFMKYPKPSDLKLGHEVPDEQLPSMWHYALHPDQMPPEP 354

Qy 344 QIYFPAFGONDKTIAEGLATFFSRGGLAKSPADLASYPDVLQANTHLOAWISFS 403

Db 355 QVYFTVFGMSDAGITNATLAFSSRHGWYEMAKYRVFLEGSFPHNDPESLNYLHTVVSF 414

Qy 404 YKGGKPYMSVYLHDFE-----APS 422

Db 415 YRKNKPYLSVYLHDFEFGQWPAFS 438

RESULT 2

C64091

adenine glycosylase - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 17-Nov-2000

C:Accession: C64091

R: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kexlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodev, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A: Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A: Reference number: A64000; MUID: 95350630; PMID: 7542800

C: Accession: C64091

A: Status: nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-378 <TIGR>

A: Cross-references: GB:U32760; GB:L42023; NID:G1573764; PIDN: AAC2418.1; PID: G1573768; T

C: Function: catalyzes the excision of adenine from a guanine-adenine mispair

C: Superfamily: A/G-specific adenine glycosylase

C: Keywords: DNA repair

Query Match 4.8%; Score 111.5; DB 2; Length 378;

Best Local Similarity 22.8%; Pred. No. 0.35;

Matches 79; Conservative 48; Mismatches 127; Indels 93; Gaps 21;

Qy 142 LEWFDHFTK-----ALVYSEBEARTLLDRDIEIPVF-----KTQKLA 180

Db 16 LAWYDFGRHLEWQNKLYGVWLVSEVMLQQQTQVATV-----IPYFERFKTPNITA 69

Qy 181 DLSPSGDIVLKY-----IYPRKSIATGTPKER-----LMFDAKAAKFGKVAPL 228

Db 70 LANASQDEVLLHWTGTYGAYARAEHLKAAQKVRDEFNGFNFTFECQWALSQVGR-STAG 128

Qy 229 AILEEFIAERAPTLHGLFLSCDLVKEPSR-----IKVCMEROLDIASIEGIWTLNRRND 285

Db 129 AILSSVLNQYPYILDGN-----VKELVARYPAVEGWSGEKKVE-----NRLWALT-EQVT 177

Qy 286 PET--LDGLDALRELHQLLPVTE-----GLCPUP-NCFYEPGTSPOBQPFIFINFTLSPKS 338

Db 178 PTTVRADFQNMMDIGAMVCMRTKPRCDLPLNIDCLAYKNTN-WEKFP-----AKKPKK 231

Qy 339 ALPEPOLYPPAFQONDKTIAEGLATFFESRG-WGGLAKSYP--ADLASYPDVLQANTH 395

Db 232 AMPEKTYTFLILSKNGKVCLEQR-----ENSLWGLGF-CFPOFEKSLHLFLAQEKVTH 286

Qy 396 LQAWISFSYKGGKPYMSVYLHDFEAFSA-----AAQEVAMCHDGHN 436

Db 287 YQWPSFR-----HTFSHEHLDIHPIAYEMESTLQVEQAN 321

RESULT 3

SYQMA

malate synthase (EC 4.1.3.2), microbody - yeast (Pichia angusta)

C:Species: Pichia angusta

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000

C:Accession: S09294

R: Bruinenberg, P.G.; Blaauw, M.; Kazemier, B.; AB, G. Yeast 6, 245-254, 1990

A: Title: Cloning and sequencing of the malate synthase gene from Hansenula polymorpha.

A: Reference number: S09294; MUID: 90273778; PMID: 2349836

A: Accession: S09294

A: Molecule type: DNA

A: Residues: 1-555 <BRU>

C: Comment: This enzyme, together with isocitrate lyase, is very important to the glyoxylate shunt.

C: Genetics: MAS

C: Superfamily: malate synthase

C: Keywords: carbon-carbon lyase; coenzyme A; glyoxylate bypass; oxo-acid-lyase; tricarbo-

Query Match 4.6%; Score 107.5; DB 2; Length 490;

Best Local Similarity 22.7%; Pred. No. 1.1;

Matches 109; Conservative 50; Mismatches 149; Indels 173; Gaps 29;

Qy 9 TLLPKPF--YVLSQALNLSNKHDTKWWYSTAPMFMATMAGAGYDVHAQYKFLCIHREVIIPAL 69

Db 59 TLLPSPFPKPVFEAVHIQ-----LFASLYHFIAY-----EFDELIDIKNVV 102

Qy 66 -----IPALGPVPEKG--QPM-----HMKSHLTRFGLPFELSN----- 97

Db 103 KTDGFTFNMVLEILKKVKAQGLKQPVLTATQSDYMCHEKQYSAEYGLK-QIENINNIASSM 161

Qy 98 --YSKSLLRPAFPLGSLTGTGKDDPPNTQAIRPVLDLQKAMVPEGLDLEWFDHFTK---A 151

A: Accession: S09294

A: Molecule type: DNA

A: Residues: 1-555 <BRU>

C: Comment: This enzyme, together with isocitrate lyase, is very important to the glyoxylate shunt.

C: Genetics: MAS

C: Superfamily: malate synthase

C: Keywords: carbon-carbon lyase; coenzyme A; glyoxylate bypass; oxo-acid-lyase; tricarbo-

Query Match 4.8%; Score 111.5; DB 1; Length 555;

Best Local Similarity 23.5%; Pred. No. 0.61;

Matches 73; Conservative 39; Mismatches 89; Indels 109; Gaps 20;

Qy 16 YVLSQALNLSNKHDT-----KWWYSTAPMFMATMAGAGYDVHAQYKFLCIHREVIIPAL 69

Db 269 YOLDEVLYALURDHSAGLNGRWY-----MFSTI-----KELRNQKHILP-- 309

Qy 70 GPYPEKQPMHWKSHLTRFGLPFELSFNYSKSLRFAFE-----PLGSLTGT---KDDP-F 121

Db 310 -----DRHQVTMTVFP--MTNYVKQLIKICHKGVHAMGGMAATIPKDDPEK 355

Qy 122 NQAIRPVLPQ-LKAMVPGDLLEWFDH---FTKALVVSSEBEART-----LLDRDIEIPVFK 173

Db 356 NAAAMEAVRQDKLEVLAGHDGTWIAHPGLLPALTALSVEQHEMPTPNQIHVQKVEI---- 411

Qy 174 TONKLAADLEPSGDIVLK-----TIVYPRKSIATGTPKERLMFPAKA----- 217

Db 412 TEADLVDTNIPDGKLTWKGVSANIYIGLNWESWRLGCG-VPINNLMEDRATAEVSRLQ 470

Qy 218 -----ADKPGKVAATPLALEEFI-----AERAPTLHGLFLSCDLVKEPSR-RI 259

Db 471 LYSWCKHAKVMDDTGKTIPT-----EFISKLIIDEAER-----CAANKPNKFKI 515

Qy 260 KVICMERQLD 269

Db 516 AADCLKKEIN 525

RESULT 4

T23791

hypothetical protein M176.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T23791

R: Wilkinson, J.

A: Submitted to the EMBL Data Library, August 1996

A: Reference number: 219799

A: Accession: T23791

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA

A: Residues: 1-490 <WIL>

A: Cross-references: EMBL: Z78412; PIDN: CAB01655.1; GSPDB: GN00020; CESP: M176.2

A: Experimental source: clone M176

C: Genetics:

A: Gene: CESP: M176.2

A: Map position: 2

A: Introns: 35/3; 259/2; 294/1; 471/3

C: Superfamily: glutathione synthase

Query Match 4.6%; Score 107.5; DB 2; Length 490;

Best Local Similarity 22.7%; Pred. No. 1.1;

Matches 109; Conservative 50; Mismatches 149; Indels 173; Gaps 29;

Qy 9 TLLPKPF--YVLSQALNLSNKHDTKWWYSTAPMFMATMAGAGYDVHAQYKFLC-IHREVI 65

Db 59 TLLPSPFPKPVFEAVHIQ-----LFASLYHFIAY-----EFDELIDIKNVV 102

Qy 66 -----IPALGPVPEKG--QPM-----HMKSHLTRFGLPFELSN----- 97

Db 103 KTDGFTFNMVLEILKKVKAQGLKQPVLTATQSDYMCHEKQYSAEYGLK-QIENINNIASSM 161

Qy 98 --YSKSLLRPAFPLGSLTGTGKDDPPNTQAIRPVLDLQKAMVPEGLDLEWFDHFTK---A 151

Db 162 GAHALRLTEWHIRVLKAL-NISDDVI--QRAIPENKPIPMIAEALFKAW-SHFSNPAUV 217
QY 152 LVVSEAEATLTD-ROIEIPVFK-----TQ--NKLA-----ADLEPSGDIVLK 191
Db 218 LVVVENVNQIDORHVEYELEKLGVPMTCIIRNLTCQYELSLNDRSLMDIGRQVAI 277
QY 192 TYI-----YPRIK-----SIATGPKERLMFDAKAADKFKGVATPLALIEE 233
Db 278 VYFRAGYSPDHYSTKEWARERVELSTAIKTPWGLQ--VANTYKTKQVLSEGVLER 334
QY 234 FIAERAPTLGLHFLSCDLVKPESRIKVKYCMERQDL-ASIEGIWTLNGRRNDPPTLD-- 290
Db 335 FIG-----KPREAR-----DIRASFAGMWAL--ENTDEVTMKV 366
QY 291 -GLDALRELWQLPVTEG-----LCPLNCEYEPGTSPOEQLPPIINFTLSP 336
Db 367 AGAQKHEAPVLKPOTEGGAULTGDEMVMQRELPE-----SEGAFILMEKLRP 417
QY 337 -----KSALPEPQIYFPAFGQND-----KTIAEGLATFPFESRGWGLAKS 376
Db 418 MIENVLVLAKKPIITPAKAVSELGVGYAFGRKDAPELKTAGHLLRTKPESTANGGVAAG 477
QY 377 Y 377
Db 478 H 478

RESULT 5

S74207

lipoxigenase (EC 1.13.11.12) - cucumber

C/Species: Cucumis sativus (cucumber)

C/Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000

C/Accession: S74207; S74137

R/Hoehne, M.; Nellen, A.; Schwennesen, K.; Kindl, H.

Eur. J. Biochem. 241, 6-11, 1996

A/Title: Lipid body lipoxigenase characterized by protein fragmentation, cDNA sequence a

A/Reference number: S74137; MUID:97054584; PMID:8898881

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-878 <HOE>

A/Cross-references: EMBL:X92890; NID:G1296511; PIDN:CAA63483.1

A/Experimental source: tissue cotyledones; clone PCSLBLOX221

A/Accession: S74137

A/Molecule type: protein

A/Residues: 196,'X',198-204

A/Superfamily: lipoxigenase

C/Keywords: oxidoreductase

Query Match 4.3%; Score 101; DB 2; Length 878;
Best Local Similarity 21.7%; Pred. No. 8.5;
Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;

QY 61 HREVIIPALQ-----PYPEK---GQPMHWKSHLTRGLPFELSFNYSKSLRFAFEPLGS 112
Db 235 HR-----PILGGTTEYPYPRGRTGRPSRDH-----NYESRL-----SPINS 273
QY 113 LT--GTKDDDFN-----TQAIRVLQDLKAMPGLDLEWDFHFTKALVVS 155
Db 274 LDIVYPKDENFGHLKMSDFLGYYTLKALSISIKPGLQSIQIDVTP---NEFDNF----- 322
QY 156 EEEARTLIDEDIEP--VFKTONKLAADLEPSGDIVLKYIYPRIKSIATGPKERLMFD 213
Db 323 -KEVDNLFEGFFIPNAPAT--LTEDLTP-----PLFKALVRNDCERKFLKP 367
QY 214 AIKAADKFGKV--ATPLALIEEPIAERAPTLGLHFLSCDLVKPESRIKVKYCMERQLDIA 271
Db 368 TPEVW-KDNKIGWSTDEEFAREMLAGNPPLII-----RRLEAFPTSKLDP- 412
QY 272 SIEGIWTLNGRRNDPPT-----LDGLDALRELWQ-----LLPVTGELCLPNCYEPPT 320
Db 413 -----NVYGNQNSTITEEHKHLGDLTVDEAMKQNRLYIVDFHDMALPYLTRMNATST 466

QY 321 SPQQLPFII---NFTLSP---KSALPEPQ-----IYFPAFGQNDKTIAEGLATFF 365
Db 467 KTYATRLLKDDGTTLKPLVLIELALPHPODQGLGALSKLYFPAENGQVQKSI----- 518
QY 366 ESRGWGLAKSYPADLASIYVPDVLQDTANHQLQAWISFYKGYKPKFVMSVYLHT----FFAFS 422
Db 519 ---W-QLAKAY-----VTVNDVGYHQLISHWLHLTHAVLPFV 551
QY 423 AAA-QEVAMCHDGH 435
Db 552 IATHRLSVLPHPIH 565

RESULT 6

AB2255

hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C/Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C/Accession: AB2255

R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, I.

J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: AB2035; MUID:20406833; PMID:10952301

A/Accession: AB2255

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1621 <HEI>

A/Cross-references: GB:AB004181; GB:AB003852; NID:9655454; PIDN:AAF94159.1; GSPDB:GN00:

A/Experimental source: serogroup O1; strain N16961; biotype El Tor

C/Genetics:

A/Gene: VC0998

A/Map position: 1

Query Match 4.3%; Score 101; DB 2; Length 1621;
Best Local Similarity 20.5%; Pred. No. 21;
Matches 67; Conservative 47; Mismatches 99; Indels 114; Gaps 16;

QY 116 TKDDPFNTQAIRPVLQDLKAMPGLDLEWDFHFTKALVSEEAARTLLDRDIEIPVKTQ 175
Db 1295 TEEDALADAQLEPAAE--SEVEPELEL-----ASVLEEDSFTELD-ELDIPYTEK 1343
QY 176 NKLA-ADLEPSGDIVLKYIYPRIKSIATGPKERLMFDAKAADKFKGVATPLALIEEF 234
Db 1344 DALADAQLEPAESEVEPELGDDETETLAQETESDALVAD-----EDLLASVESA 1392
QY 235 IAERAPTLGLHFLSCDLVKPESRI-----KVCMERQLDLASIE 274
Db 1393 VDEVPPELLG---ATQVVPPTQSLANKAPDEEALHDLSDNPQGEKPFSDRPLDAKTI- 1448
QY 275 GIWTLNGRRNDPPTLDGLDALREL----WQLLPVTEGLCLPNCYFBFGTSPQQLPFII 330
Db 1449 -----DSAGMD-IDAMLQMGEDWNGFHLT-----PDQQAQLPDDV 1483
QY 331 NFTLSPK-----SALPEQIYFPAFGQNDKTIAEGLATFPFESRGWGLAKSYPADLAS 383
Db 1484 -----PEDEQAIWASETPEQA-----KPENWGS-----QEDLLD 1513
QY 384 YYPDVD-LQTANHQLQAWISFYKGYKGP 409
Db 1514 FDPQRDGYMTIDELWAQVESEEQGNP 1540

RESULT 7

G87379

heptosyltransferase family protein [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 30-Sep-2002

C/Accession: G87379

R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: G87379
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-323 <STO>
A/Cross-references: GB:AE005673; NID:gl3422349; PIDN:AAK23035.1; GSPDB:GN00148
C/Genetics:
A/Gene: CCL051
C/Superfamily: ADP-heptose-LPS heptosyltransferase II

Query Match 4.3%; Score 100; DB 2; Length 323;
Best Local Similarity 19.8%; Pred. No. 2.3;
Matches 53; Conservative 33; Mismatches 90; Indels 92; Gaps 10;

QY 178 LADLPSSGDIVLK-----TYIYRIK-STATGTPKERL-----MFAIFA---ADKF 221
DB 9 LVTKLGLGDFVLALAMRKIRBAHPKAKITLLTPFFALAKLSYFNSVETDGRPDDF 68
QY 222 GKVATPLAILEEETIARA-----PTLLGHFLSCDLVSPSRIRKV 261
DB 69 GQTLAMIGRKARYDEVYDLQTSRTNGYFQMLRPPFQWNGIAGVCCSLPQKGSARLTM 128
QY 262 YCWERQLDLASIEGIWTLNGRRNDPDTLGLDALRELWQLLPVTEGLCPPLNCFPVPGTS 321
DB 129 HTLERQADQLRAAGIW-----PPDAPTEPGSA 154
QY 322 PQQLPEINFTLSBK-----SALPEQIYFPAGQNDKTIAGLATFFESRCG----- 370
DB 155 PPDLGLSWILKHKHPRVAGAAAPRYLP-----VPGSAHLEKR-WPVECTAQ 204
QY 371 -GGLAKSPADLASYYPDVLDLOTANHLQ 397
DB 205 LGSLLAKGLDIVIIGQPQESAMARIHQ 232

RESULT 8

JC4240
lanosterol 14alpha-demethylase (EC 1.14.14.-) cytochrome P450 51 - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 21-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 28-Jul-2000
A/Accession: JC4240; PC4069
R/Sloane, D.L.; So, O.Y.; Leung, R.; Scarafia, L.E.; Saidou, N.; Jarnagin, K.; Swinney, Gene 151, 243-248, 1995
A/Title: Cloning and functional expression of the cDNA encoding rat lanosterol 14-alpha
A/Reference number: JC4240; MUID:95394364; PMID:7665087
A/Accession: JC4240
A/Molecule type: mRNA
A/Residues: 1-486 <SLO>
A/Cross-references: GB:U17697; NID:G699395; PIDN:AAA87074.1; PID:G699396
A/Accession: PC4069
A/Molecule type: protein
A/Residues: 43-60; 206-213; 254-267; 369-387; 420-428; 433-443 <SL2>
A/Experimental source: liver
C/Comment: This enzyme is a cytochrome P-450 enzyme which catalyzes one of the essential
C/Superfamily: human cytochrome P450 CYP51; cytochrome P450 homology
C/Keywords: chromoprotein; heme; iron; liver; metalloprotein; monooxygenase; oxidoreduct
F;251-454/Domain: cytochrome P450 homology <P45>
F;432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 4.3%; Score 99.5; DB 2; Length 486;
Best Local Similarity 19.3%; Pred. No. 4.7;
Matches 92; Conservative 62; Mismatches 190; Indels 133; Gaps 19;

QY 39 MFATMAGAGYDVAQYKP-LCHRSVITPAGLPYEPK-GQPMHWKSHLTRG-LPFFLS 95
DB 9 LUSTLIACFTLSLVFLVAGHVMQPLPAGAKSPFY-TSYPFPLGHAIAGKPIBFL 68
QY 96 FN-YKSLRLFAPEPLGS-----LTGKDDPFNTQAI-----RPV----- 129
DB 69 ENAYEKVGFVSTMGKFTYLLGSDAAALLFNSKEDLNAAEVYGRLLTFVFGKWAY 128

QY 130 -----LQDLKAMVPGDLLEWFDHFTKALVGBEEAR----- 160
DB 129 DVPNAVLEQKILKSGNIA--HFQYVSIIEKEAKYFKSWGSEGERNVFEALSELI 185
QY 161 -----TLDRDIEIPVKTOKNLAADLE-----P8-----GDVLTKT 192
DB 186 ILTASHCHLHGKEIRSQLNEKVAQIYADLDGGFSHAAMLPGWLPFLPSFRDRRAHREIKN 245
QY 193 YIYPRY-KSTATGTPKERLMFDAIKADKFGKATP-----LAILLEFIAERAPTL 242
DB 246 IFYKAIQRRLSKPEADIIQLTLDSTYKDGRELTDDIAGMLIGLLAGCHTSSTTSW 305
QY 243 LGHFLSCDLVKPSESRI---KVVC-----MERQLDLASIE-----GIWTL 279
DB 306 MGFFLARD--KFLQDKCYLQKTVCGSDLPPLVTEGLKDLNLLDRCIKETRLRPPIMTM 363
QY 280 NGRNDEPDTLGLDALRELWQLLPVTEGLCPPLNCFPVPGTSPOBQLPFIINFTLSPKSA 339
DB 364 MRMAKTPTOTVAG-----YTIPPGHQVCVSPVQRLKDSWVERLDNFDPDYQNPA 415
QY 340 LPEQIYFPAGQNDKTIAGLATFFESRCGWGGLAKSPADLAS-YYPDVDLOTANH 395
DB 416 SGEXFAVYFPGAGHRCIGENFAVQIKTIWMTLRIYEFDLINGYFVSVNYTMIH 472

RESULT 9

JC4758
lanosterol 14alpha-demethylase (EC 1.14.14.-) cytochrome P450 51 - rat
A/Alternate names: pRT-9 protein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 28-Jul-2000
A/Accession: JC4758; PC4170; JC5888; JC2334
R/Aoyama, Y.; Noshiro, M.; Gotoh, O.; Imaoka, S.; Funae, Y.; Kurosawa, N.; Horiuchi, T.; J. Biochem. 119, 926-933, 1996
A/Title: Sterol 14-demethylase P450 (P45014DM) is one of the most ancient and conserved
A/Reference number: JC4758; MUID:96389999; PMID:8797093
A/Accession: JC4758
A/Molecule type: mRNA
A/Residues: 1-503 <AOY1>
A/Cross-references: DDBJ:DS5681; DDBJ:D29962; NID:gl020094; PIDN:BAA09529.1; PID:G870754
A/Accession: PC4170
A/Molecule type: protein
A/Residues: 52-109; 305-319; 324-334; 467-482; 488-499 <AOY2>
A/Experimental source: liver
R/Noshiro, M.; Aoyama, Y.; Kawamoto, T.; Gotoh, O.; Horiuchi, T.; Yoshida, Y. J. Biochem. 122, 1114-1121, 1997
A/Title: Structural and evolutionary studies on sterol 14-demethylase P450 (CYP51), the
A/Reference number: JC5888; MUID:98158318; PMID:9498553
A/Accession: JC5888
A/Molecule type: DNA
A/Residues: 1-503 <NOS>
A/Cross-references: DDBJ:AB004087
R/Aoyama, Y.; Funae, Y.; Noshiro, M.; Horiuchi, T.; Yoshida, Y. Biochem. Biophys. Res. Commun. 201, 1320-1326, 1994
A/Title: Occurrence of a P450 showing high homology to yeast lanosterol 14-demethylase (

A/Reference number: JC2334; MUID:94296405; PMID:8024575
A/Accession: JC2334
A/Molecule type: mRNA
A/Residues: 74-503 <AOY>
A/Cross-references: DDBJ:D29962
A/Experimental source: liver
C/Genetics:
A/Gene: CYP51
A/Introns: 59/3; 91/3; 150/3; 193/1; 251/2; 291/2; 356/3; 388/3; 445/1
C/Function:
A/Description: catalyzes the removal of the 14-methyl group of 14-methylsterols
C/Superfamily: human cytochrome P450 CYP51; cytochrome P450 homology
C/Keywords: chromoprotein; heme; iron; liver; metalloprotein; monooxygenase; oxidoreduct
F;308-471/Domain: cytochrome P450 homology <P45>
F;445/Binding site: heme iron (Cys) (axial ligand) #status predicted
Query Match 4.3%; Score 99.5; DB 2; Length 503;
Best Local Similarity 19.3%; Pred. No. 4.9;

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70922

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-344 <COL>

A;Cross-references: GB:295150; GB:AL123456; NID:g2250708; PID:CA080361.1; PID:g2076700

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: RV3127

C;Superfamily: Mycobacterium hypothetical protein RV3131

Query Match 4.2%; Score 97.5; DB 2; Length 344;
Best Local Similarity 20.7%; Pred. No. 4.1;
Matches 65; Conservative 38; Mismatches 94; Indels 117; Gaps 14;

QY 80 HKSHLTFGLPELSFNYSKSLRFAFPLGLSLTGT-----DDPFN 122

DB 73 HWQANITRFPP-----NQPDQATVEFSPIDHTVAGQRAQAILQRTDRLPDSMY 127

QY 123 TQAIRPVLD-LKAMPVGLDLEWDFHTKALVSE-----156

DB 128 WHLFEALRDVADKDVAMLDVSDQDRLVVASQLSEVLRRDPYHAELEWWTSPFVL 187

QY 157 -----EARTLLDRDIEIPVFTONKLAADLEPSGDIVLKYIIPRIKSIAT 203

DB 188 AHGVPPDTLASDAERLURVDLGRD--PVRSYQNRRA---ELADD-----RSKVLVL 233

QY 204 GTPKERLMFDAIKADKFGKATPLAILBEFIAERATLLGHFLSCDLVKPSESRIKYVC 263

DB 234 STPSS-TRADALRC---GVLSTI-LLECTMAGMATCTLTH-----LISSDSRDIRVG 282

QY 264 MERQDLASIEGTWTNGRNDPETHDGLDALRELWQLPVTGELCPNCFYEGTSPQ 323

DB 283 LTRQ-----RGEQAL-----IRVGTIA--PPLAAVPAFPTPR 311

QY 324 EQLPFIINFTLSPK 337

DB 312 RPLDSVLQIRQTPB 325

RESULT 13

A71221

C;Species: Pyrococcus horikoshii

C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 03-Nov-2000

C;Accession: A71221

R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Accession: A71221

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-621 <RAW>

A;Cross-references: GB:AP000001; NID:g2336128; PID:BAA29096.1; PID:g3256413

A;Experimental source: strain OT3

A;Note: this accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics:

A;Gene: PH0028

C;Superfamily: probable aldehyde ferredoxin oxidoreductase aor-4

C;Keywords: oxidoreductase

Query Match 4.2%; Score 96.5; DB 2; Length 621;
Best Local Similarity 19.2%; Pred. No. 12;
Matches 59; Conservative 31; Mismatches 83; Indels 135; Gaps 13;

QY 182 LEPSPGDIVLK-----TVIYPRIKSIAT--GTPKERLMFDAIKADKFGKATPLAILEFI 235

DB 56 LSPGNKIVFAPGLTGLIPSSKVITVSKSPETRLITDS--SGDAFG-----101

QY 236 AERAPTLGHFLSCDLVKPSESRIKYCMEROLDIASIEGTWTNGRNDPETHDGLDAL 295

DB 102 ----PKLKGHFDALIEGKSEEPVLYIHGGVDILPAGELW---GKGN-----YETA 147

QY 296 RELHQ-----LLPVTEGLCLPLNCFY-----EGETS 321

DB 148 RELMKYFEASIASIGPAGERLVRIANIYYDTORASRGGLGAVMGSKLKAIVVEPEGK 207

QY 322 PQEQLP-----FINTFLSPKSALPEFQIYFFAFGQNDKTIAREGLATFFESRGWG- 371

DB 208 PEVANPEFEALWNEFYERFSTDPK-----YEHSRYNGT 241

QY 372 -----GLAKSY-----PADLASYYPDVDTQANHLQAW-----I 400

DB 242 TDGLRSSASLGMSPAYNFSRPYIPEELASKLAGDEVKKYEVPEWYIHGKSCPIKCARYI 301

QY 401 SFSYKGGK 408

DB 302 EYEYKGRK 309

RESULT 14

S69079

N;Alternate names: hypothetical protein YPR097w - yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

C;Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 19-Apr-2002

C;Accession: S69079

R;Couch, J.

A;Description: The sequence of S. cerevisiae cosmid 9513.

A;Reference number: S69057

A;Accession: S69079

A;Molecule type: DNA

A;Residues: 1-1073 <COU>

A;Cross-references: EMBL:U51033; NID:g1230676; PID:g1230699; GSPDB:GN000016; MIPS:YPR097w

C;Genetics:

A;Gene: MIPS:YPR097w

A;Cross-references: SGD:S0006301

A;Map position: 16R

Query Match 4.1%; Score 95.5; DB 2; Length 1073;

Best Local Similarity 19.5%; Pred. No. 32;

Matches 66; Conservative 61; Mismatches 116; Indels 95; Gaps 17;

QY 9 TLLPKPPYVLSQLNLSNKDHTK-----WW--YSTAPMFATMMAG-AGYDVH 52

DB 556 SLKDEKYLMLLVEI--KNTKVEDLSPLLQDFVWCKIYISSMYQMFGLNDNSYEL 613

QY 53 AQYKFLCIHREVIIPALGPYPERGQPMHMKSHLTRFLPPELSFNYSKSLRFAFPPLGS 112

DB 614 TQIRRL--HK-----LMPYVMGQIMKFTNPIAINRGMIEL-----FMAQPF 655

QY 113 LTGTDKDPFNTQAIRPVLDLKAIVGLDLEWDFHTKALVSEEBEARTLLDRDILP 172

DB 656 -----HSLLOTFSTILTDDLK-----TQKVAIKE-----LEKKI----- 695

QY 173 KTONKLAADLEPSGDIV--LKYIYIPRIKSIATGTPKERLMF--DAIKAADKFGKATP 227

DB 686 -----AENDPCASVVTKLDFVF-----NNDTKDEHDTKLTMDAVNAESNMNVP 734

QY 228 LAILEPIA-----ERAPTLGHFLSCDLVKPSESRIKYCMERO-LDLASIEGTWTN 280

DB 735 LIVLMKSAANLIPDEVVAGLIESYSSWKLQKEDTDALANTSEDQSGIYFTHYKDLW 794

QY 281 GRNDPETHDGLDALRELWQLPVTGELCPNCFYEP 318

DB 795 IKEHDQ-----LMQLAQDDELQMLKAIVTIYEP 826

RESULT 15

GNVWTR

genome polyprotein 2 - tomato ringspot virus (strain raspberry)

N;Contains: coat protein

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jun-2000

C/Accession: JQ1093
R/Rott, M.E.; Tremaine, J.H.; Rochon, D.M.
J. Gen. Virol. 72, 1505-1514, 1991
A/Title: Nucleotide sequence of tomato ringspot virus RNA-2.
A/Reference number: JQ1093; MUID:91311402; PMID:1856689
A/Accession: JQ1093
A/Molecule type: genomic RNA
A/Residues: 1-1882 <ROT>
A/Cross-references: GB:D12477; GB:D01129; NID:G222674; PIDN:BAA02043.1; PID:G222675
A/Note: it is uncertain whether Met-1 or Met-122 is the initiator
C/Genetics:

A/Map position: segment 2
C/Superfamily: tomato ringspot virus genome polyprotein
C/Superfamily: coat protein; glycoprotein; polyprotein
F/1321-1882/Product: coat protein #status predicted <MAT>
F/269,295,1183,1316,1543,1561,1735/Binding site: carbohydrate (Asn) (covalent) #status F

Query Match 4.1%; Score 95.5; DB 1; Length 1882;
Best Local Similarity 20.3%; Pred. No. 72;
Matches 84; Conservative 50; Mismatches 129; Indels 151; Gaps 24;
QY 33 WYSTAMPFATMAGAGYDVAQYKFLC-----IHREVIIP-----ALGPYPE 74
DB 608 WSSPLFLFANFKVNRG-----ACFLQVLQFQVVLPDECMDDLLEDFEDQLPEGPPLPS 658
QY 75 KGQPMHWKSHLTRFGLPFELSFNYSK-----SLRFAFEPLGSLTG 115
DB 659 ----FSWSS-----PLPLFASPKVNRGACFLQVLPARKVVSDBFMDVLPFLFSP--VSH 707
QY 116 TKDDPNTQAIRVLQDL-----KAVPGLDLE-WPDHFTKALVSEEEARTLLDRDIRP 170
DB 708 QEEPEMVPVLEAADSVDGITEAFEDDLECESFYDSYS-----DEEEAEW-----AEVP 757
QY 171 VFKTOKLAADLEPSGDIVLKTIVYPRIKSIATGTPKERLMFDAIKADKFGKVATPLAI 230
DB 758 RCTMBELCASLTLAD-----AEGLRKSHGVF-----LKELVTYLOS 795
QY 231 LEE-FIAERAPTLHGHLSCDLVKP-----SESRIKVCMERQLD-----LASIGIWT 279
DB 796 FEESPLYSSRA-----FYSVK-VKPVYRKKFECHIDCTCLDGNMGEWRESVDAMMRC 848
QY 280 NGR-----RNDPETLDGL-----DALRELWQLLPVTE-----GLCPLPN 313
DB 849 PGLLNTKRTFTDDWERYOYLRIGNEGRYRNWRVNLNLEMDLSLHYPFESSAPVOS 908
QY 314 CFY-----EPGTSPEQLPFIINFTLSPKSGALPEPFIYFPAFGQNDKXTIAGLAT 363
DB 909 SLFSRVVDRCATLASSIPFVTRNC--QSSLGTP-----GLNVHTIHQEAPT 953

Search completed: July 22, 2004, 18:46:29

Job time : 21 secs

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:34:31 ; Search time 13 Seconds
(without alignments)
1750.358 Million cell updates/sec

Title: US-10-099-704-2
Perfect score: 2322
Sequence: 1 MEISKKAATLLPKPFVLSQ.....FEAFSAQAQEVAMCHDGHNP 437

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	111.5	4.8	378	1 MUTY_HAEIN	P44320 haemophilus
2	111.5	4.8	555	1 MASY_PICAN	P21360 pichia angu
3	110.5	4.8	462	1 ACDS_METKA	Q8txf1 methanopyru
4	99.5	4.3	503	1 CPS1_RAT	Q9byk8 rattus norv
5	98.5	4.2	2649	1 P285_HUMAN	P25247 tomato ring
6	95.5	4.1	1882	1 POL2_TRSVR	P49194 mus musculu
7	95	4.1	610	1 IRBP_MOUSE	Q9z7p1 chlamydia p
8	94.5	4.1	430	1 SYH_CHLPN	P03362 human t-cel
9	94.5	4.1	896	1 POL_HILLA	Q55522 synechocyst
10	94	4.0	910	1 SYV_SYNY3	Q978s0 streptococ
11	93	4.0	833	1 SYL_STRPN	P27282 venezuelan
12	93	4.0	2492	1 POLN_EEVT	P33744 clostridium
13	92.5	4.0	862	1 ADHE_CLOAB	P34453 caenorhabdi
14	92	4.0	282	1 YMA8_CABEL	Q9z1w1 helicobacte
15	92	4.0	443	1 HSLU_HELPJ	Q91lp5 arabidopsis
16	92	4.0	500	1 C72W_ARATH	Q09785 schizosacch
17	92	4.0	1017	1 GCSP_SCHPO	P36327 venezuelan
18	92	4.0	2485	1 POLN_EEVV3	P41712 bombyx mori
19	91.5	3.9	466	1 NMT_SCHPO	Q8bj37 mus musculu
20	91.5	3.9	986	1 DPOL_NPVBM	Q985w2 oncorhynch
21	91	3.9	609	1 TYDP_MOUSE	Q8dbb6 streptococ
22	91	3.9	766	1 HIFA_ONCMY	Q58477 methanococ
23	91	3.9	833	1 SYL_STRR6	O52225 thermus fil
24	90.5	3.9	521	1 SYS_METJA	P41410 schizosacch
25	90.5	3.9	833	1 DPOI_THEFI	P34756 saccharomyc
26	90.5	3.9	852	1 RA54_SCHPO	P96142 thermus the
27	90.5	3.9	2278	1 FABI_YEAST	P36328 venezuelan
28	90	3.9	862	1 SYV_THETH	P98161 homo sapien
29	90	3.9	2492	1 POLN_EEVV3	P36122 saccharomyc
30	90	3.9	4303	1 PKDI_HUMAN	Q12558 aspergillus
31	89	3.8	765	1 YK07_YEAST	P404899 mus musculu
32	89	3.8	985	1 AGLU_ASPOR	
33	88.5	3.8	451	1 KPT3_MOUSE	

RESULT 1

MUTY_HAEIN

ID MUTY_HAEIN STANDARD; PRT; 378 AA.

AC P44320;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE A/G-specific adenine glycosylase (EC 3.2.2.-).

OS MUTY OR H10759.

GN Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Rd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae

Rd.";

CC -!- FUNCTION: Adenine glycosylase active on G-A and C-A mispairs (By

similarity).

CC -!- COPACTOR: Binds a 4Fe-4S cluster which is not important for the

catalytic activity, but which is probably involved in the proper

positioning of the enzyme along the DNA strand (By similarity).

CC -!- SIMILARITY: Belongs to the nth/mutY family.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; U32760; AAC22418.1; -.

DR PIR; C64091; C64091.

DR HSRP; P17802; IMUY.

DR TIGR; H10759; -.

DR InterPro; IPR003265; Endo_3c.

DR InterPro; IPR004035; EndoIII_FCL.

DR InterPro; IPR004036; EndoIII_HhH.

DR InterPro; IPR003651; FeS_bind.

DR InterPro; IPR000445; HHH.

DR InterPro; IPR005760; MutY.

DR Pfam; PF00730; HHH-GPD; 1.

DR Pfam; PF00633; HHH; 1.

Q36552 homo sapien
Q8p726 xanthomonas
Q04264 saccharomyc
Q12767 homo sapien
P38931 saccharomyc
Q64459 mus musculu
O57606 aquifex aeo
P09278 varicella-z
Q57576 methanococc
P27237 salmonella
P30313 thermus the
P39812 bacillus su

DR SMART; SMO0478; ENDO3c; 1.
 DR SMART; SMO0525; FES; 1.
 DR TIGRAME; TIGR01084; muty; 1.
 DR PROSITE; PS00784; ENDONUCLEASE_III_1; 1.
 DR PROSITE; PS01155; ENDONUCLEASE_III_2; 1.
 DR DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;
 KW Complete proteome.
 FT METAL 197 197 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 204 204 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 207 207 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 213 213 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 213 213 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 378 AA; 43505 MW; C43F625131B2A21 CRC64;
 Query Match 4.8%; Score 111.5; DB 1; Length 378;
 Best Local Similarity 22.8%; Pred. No. 0.14;
 Matches 79; Conservative 48; Mismatches 127; Indels 93; Gaps 21;
 QY 142 LEWFDHTK-----ALVSVSEEARLTLDRDIEIPV-----KTQNKLA 180
 DB 16 LAWYDFGRKHLFPWQNKTLTYGVMLSEVLMQQTQVATV-----IPYPERFIKTFPNITA 69
 QY 181 DLSPSGDIVLKY-----IYPRIKSIATGTPK-----LMFADKAADKFGKVATPL 228
 DB 70 LANASQDEVILHLWTGLGYARARNLHKAQKVRDEPNFGNPTNFQVWALSGVGR-STAG 128
 QY 229 AILEEPIAERAPLLGHFSLCDIVKPSER---IKVCMERQOLDIASIEGIWTLNRRND 285
 DB 129 AILLSVINQPYPLDGN-----VKEVLARYFAVEGWSGEKKVE-----NRLWALT-EQVT 177
 QY 286 PET--LDGLDALRELWQLLPVTE---GLCPLE-NCFYEPGTSPOBOLFIINFILSPKS 338
 DB 178 PTRVADFQNMMDIGAMVCMRTKPCDLCPLNIDCLAYKTN-WEKFP-----AKPKK 231
 QY 339 ALPEQIYPPAFQNDKTIAGLAEIATFFESRG-WGLAKSY-ADLASYPVDVDTQANH 395
 DB 232 AMPEKTYTFLILSKNGKVCLEQR---ENSLWGLGF-CFPOFEDKSSLHLHLAQEKVTH 286
 QY 396 LQAWISFSYKPKPYMSVLVHFEAPSA-----AAQEVAMCHDGN 436
 DB 287 YQWPSFR-----HTFSPHLDIHIPIYAEEMESTLVCVOAN 321
 RESULT 2
 ID MASY_PICAN STANDARD; PRT; 555 AA.
 AC P21360;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Malate synthase, glyoxysomal (EC 2.3.3.9).
 GN MAS.
 OS Pichia angusta (Yeast) (Hansenula polymorpha).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4905;
 RN [1]
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 234-253.
 RC STRAIN=MAYA CB54732;
 RX MEDLINE=90273778; PubMed=2349836;
 RA Bruinenberg P.G., Blaauw M., Kazemier B., Ab G.;
 RT "Cloning and sequencing of the malate synthase gene from Hansenula
 RT polymorpha";
 RL Yeast 6:245-254(1990).
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CC CoA.
 CC -!- PATHWAY: Glyoxylate bypass; second step.
 CC -!- SUBCELLULAR LOCATION: Glyoxysomal.
 CC -!- INDUCTION: By ethanol.
 CC -!- SIMILARITY: Belongs to the malate synthase family.
 DR PIR; S09294; SYHQA.
 DR InterPro; IPR006252; Malate synthase.
 DR InterPro; IPR001465; Malate synthase.

DR TIGRAME; TIGR01344; malate syn A; 1.
 DR PROSITE; PS00510; MALATE SYNTHASE; 1.
 KW Transferase; Glyoxylate Bypass; Tricarboxylic acid cycle; Glyoxysome.
 SQ SEQUENCE 555 AA; 63255 MW; 88137B7B2E3E303 CRC64;
 Query Match 4.8%; Score 111.5; DB 1; Length 555;
 Best Local Similarity 23.5%; Pred. No. 0.24;
 Matches 73; Conservative 39; Mismatches 89; Indels 109; Gaps 20;
 QY 16 YVLSQALNLNKHDT-----KMWSTAPMFATMAGAGYDVHAQYKFLCIHREVIIPAL 69
 DB 269 YQDEVY-YALDRHSAGLNGRWY---NFSII-----XRLNQKHLP-- 309
 QY 70 GPYPEKQPMHWKSHLTFGLPFELSFNYSKSLRFAFE---PLGSLTGT---KDDP-F 121
 DB 310 -----DRHQVTMTVP--MTNYVKLIKIKCHKRGVHAMGMAATPIKDDPEK 355
 QY 122 NTQAIRPVLOD-LKAMVPGLDLEWFDH---FTKALVYSEBEAET---LLDRDIEIPVK 173
 DB 356 NAAAMEAVROCKUREVLAGHDGTWIAHPGLLTALSVFQEHMPTNQIHVQKNVEI--- 411
 QY 174 TQNKLAADLPSPSGDIVLK-----TYIYPRIKSIATGTPKERIMFADAICA----- 217
 DB 412 TEADLVDTNPDGKITMKGVSANIYIGLNYMESWLRGLGC-VPINNLMEDAATAEVSRLQ 470
 QY 218 -----ADKFGKATPLALEEPI-----ABRAPTLGHFLSCDLVPSSES-RI 259
 DB 471 LYSCKHAKVMDDTGKTIPT-----EFISKLIDEEAER-----CAANKPNKFKI 515
 QY 260 KVMCMEROLD 269
 DB 516 AADCLKKEIN 525
 RESULT 3
 ID ACDG_METKA STANDARD; PRT; 462 AA.
 AC Q8TXE1;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Acetyl-CoA decarboxylase/synthase complex gamma subunit (EC 2.1.1.-)
 DE (ACDS complex gamma subunit) (ACDS complex methyltransferase)
 DE (Corrinoid/iron-sulfur component large subunit).
 GN CDHE OR MK0723.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 OC Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezhevaeva K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozhavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 RT and monophyly of archaeal methanogens";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 CC -!- FUNCTION: Part of a complex that catalyzes the reversible cleavage
 CC of acetyl-CoA, allowing autotrophic growth from CO(2) (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: Co(1) + CH(3)-H(4)SPT + H(1) = Co(III)-CH(3) +
 CC H(4)SPT.
 CC -!- COFACTOR: Binds 1 4Fe-4S cluster and factor III (Probable).
 CC -!- SUBUNIT: Heterodimer of delta and gamma chains. The ACDS complex
 CC is made up of alpha, epsilon, beta, gamma and delta chains with a
 CC probable stoichiometry of (alpha(2)epsilon(2))(4)-beta(8)-
 CC (gamma(1)delta(1))(8) (Potential).
 CC -!- SIMILARITY: The iron-sulfur centers are similar to those of
 CC bacterial-type 4Fe-4S ferredoxins

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF010365; AF01937.1; --
CC HAMAP; MF_01136; ; 1. --
CC InterPro; IPR004486; CdhD.
CC InterPro; IPR007202; FeS.
CC Pfam; PF03599; CdhD; 1.
CC Pfam; PF04060; FeS; 1.
CC PROSITE; PS00198; 4F84S FERREDOXIN; FALSE NEG.
KW Transferrase; Methyltransferase; Metal-binding; Cobalt; Iron;
KW Iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 18 18 IRON-SULFUR (4FE-4S) (PROBABLE).
FT METAL 21 21 IRON-SULFUR (4FE-4S) (PROBABLE).
FT METAL 26 26 IRON-SULFUR (4FE-4S) (PROBABLE).
FT METAL 43 43 IRON-SULFUR (4FE-4S) (PROBABLE).
SQ SEQUENCE 462 AA; 50852 MW; 1E97B54407368D3B CRC64;

Query Match 4.8%; Score 110.5; DB 1; Length 462;
Best Local Similarity 27.0%; Pred. No. 0.22;
Matches 62; Conservative 28; Mismatches 91; Indels 49; Gaps 12;

QY 125 AIRPVLQDKAMVPGLDLEWDFHFTKALVSVSEEARLT--LDRDIEIPVFTKQNKLAADL 182
DB 159 AVALLITDDPKVWEAGLDV--FDRPLLYPATENVEDLAKLAADGDCPL-----GLHARDV 212
QY 183 EPGSDIVLTYTYPRIKSIATGTPKRLMFDKAADKFGKVKATPLALEEPIARAPTL 242
DB 213 EDLVPLVVEAQQYTDLLLDLDPGT--EFGPHDVVSTTKLAEIRK--AAIEEFESFGYPTL 268
QY 243 LG----HFLSCDLVKPS--ESRIKVMERQDLASIEGIVTLNGRRNDPETLDGLDALR 296
DB 269 VTFPFALEDDPKVKAARESYLASCVLRADILIMDV----- 308
QY 297 ELWQLLPV-TEGLC-----ELPNCFYEPGT-----SPOEQLPFTI--NFTLS 335
DB 309 EPWALLPVLTORCVYTDPREQVEBGLYRIGDPDENSEVLVTNFTLT 358

RESULT 4
CP51_RAT STANDARD; PRT; 503 AA.
AC Q64654; Q64549;
AT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 51A1 (EC 1.14.13.70) (CYP11) (P450LI) (Sterol 14-alpha demethylase) (Lanosterol 14-alpha demethylase) (LDM) (P450-14DM).
GN CYP51A1 OR CYP51.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RC MEDLINE=94296405; PubMed=8024575;
RX Aoyama Y., Funae Y., Noshiro M., Horiuchi T., Yoshida Y.;
RT "Occurrence of a P450 showing high homology to yeast lanosterol 14-demethylase (P450(14DM)) in the rat liver.";
RL Biochem. Biophys. Res. Commun. 201:1320-1326 (1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Wistar;
RC MEDLINE=98158318; PubMed=9498553;
RX Noshiro M., Aoyama Y., Kawamoto T., Gotoh O., Horiuchi T., Yoshida Y.;
RT "Structural and evolutionary studies on sterol 14-demethylase P450 (CYP51), the most conserved P450 monooxygenase: I. Structural analyses

RT of the Gene and multiple sizes of mRNA.";
RL J. Biochem. 122:1114-1121 (1997).
RN [3]
RP SEQUENCE OF 18-503 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95394364; PubMed=7665087;
RA Sloane D.L., So O.Y., Leung R., Scarafia L.E., Saldou N., Jarnagin K.,
RA Swinney D.C.;
RT "Cloning and functional expression of the cDNA encoding rat
RT lanosterol 14-alpha demethylase.";
RL Gene 161:243-248 (1995).
CC -!- FUNCTION: Catalyzes C14-demethylation of lanosterol; it transforms
CC lanosterol into 4,4'-dimethyl cholesterol-8,14,24-triene-3-beta-ol.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- CATALYTIC ACTIVITY: Obtusifoliol + 3 O(2) + 3 NADPH = 4-alpha-
CC methyl-5-alpha-ergosta-8,14,24(28)-trien-3-beta-ol + formate + 3
CC NADP(+) + 3 H(2)O.
CC -!- PATHWAY: Cholesterol biosynthesis.
CC -!- SUBCELLULAR LOCATION: Microsomal (Potential).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D55681; BAA09529.1; --
DR EMBL; AB004096; BAA20354.1; --
DR EMBL; AB004087; BAA20354.1; JOINED.
DR EMBL; AB004088; BAA20354.1; JOINED.
DR EMBL; AB004089; BAA20354.1; JOINED.
DR EMBL; AB004090; BAA20354.1; JOINED.
DR EMBL; AB004091; BAA20354.1; JOINED.
DR EMBL; AB004092; BAA20354.1; JOINED.
DR EMBL; AB004093; BAA20354.1; JOINED.
DR EMBL; AB004094; BAA20354.1; JOINED.
DR EMBL; AB004095; BAA20354.1; JOINED.
DR EMBL; U17697; AAA87074.1; --
DR PIR; JC4240; JC4240.
DR PIR; JC4758; JC4758.
DR HSP; P77901; L89X.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
KW Heme; Cholesterol biosynthesis; NADP
FT TRANSMEM 24 44 POTENTIAL.
FT METAL 449 449 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT CONFLICT 181 181 E -> K (IN REF. 3).
SQ SEQUENCE 503 AA; 56706 MW; 33D8F345FF9CF21 CRC64;

Query Match 4.3%; Score 99.5; DB 1; Length 503;
Best Local Similarity 19.3%; Pred. No. 2;
Matches 92; Conservative 62; Mismatches 190; Indels 133; Gaps 19;

QY 39 MFATMAGAGYDVHQAQYKFLCIHREVIIPALGPYPEK-GQPMHWKSHLTFGLPPELS 95
DB 26 LLSTLLIACATSLVYLFRLAVGHMVQLPAGAKSPPIYSPIPFLGHAFAGKSPFEL 85
QY 96 FN-YSKSLRLRAFEPLGS-----LTGKDDPNTQAI-----RPV----- 129
DB 86 ENAYEKGPFVFSFTWVGKFTTYLLGSDAAALLFNKKNEDLNAEVYGLRTTPVFGKVAY 145
QY 130 -----LQDLKAMVPGLDLEWDFHFTKALVSVSEEAR----- 160
DB 146 DVPNAVFLQKILKSLGNIA--HFQYVSIIEKEAKEYFKSWGSGERNVFEALSELI 202
QY 161 -----TLLDRDIEIPVFTKQNKLAADL-----PS-----GDIVLKT 192


```
DR EMBL; AK055611; BAB70969.1; ALT_INIT.  
DR InterPro; IPR001900; Ribonuclease_II.  
DR InterPro; IPR007087; Znf_C2H2.  
DR Pfam; PF00773; RNB; 1.  
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.  
DR PROSITE; PS01157; ZINC_FINGER_C2H2_2; FALSE NEG.  
DR PROSITE; PS01175; RIBONUCLEASE_II; UNKNOWN_1.  
KW Transcription regulation; Activator; Nuclear protein; ATP-binding;  
KW Helicase; DNA-binding; Metal-binding; Zinc; Repeat; Zinc-finger;  
KW Alternative splicing; Polymorphism.  
DR ZN_FING 40 64 C2H2-TYPE (ATYPICAL).  
FT NP_BIND 544 551 ATP (POTENTIAL).  
FT SITE 667 670 DEAA BOX.  
FT NP_BIND 2174 2181 ATP (POTENTIAL).  
FT DOMAIN 968 1039 ALA-RICH.  
FT SITE 1075 1079 LXXLL MOTIF 1.  
FT SITE 1118 1122 LXXLL MOTIF 2.  
FT SITE 1173 1177 LXXLL MOTIF 3.  
FT SITE 2012 2016 LXXLL MOTIF 4.  
FT SITE 2229 2233 LXXLL MOTIF 5.  
FT VARSPLIC 1 569 Missing (in isoform 2).  
FT VARSPLIC 570 576 /FTid=VSP_007297.  
FT VARSPLIC 836 863 LICHITN -> MSSSPSR (in isoform 2).  
FT VARSPLIC 2163 2181 Missing (in isoform 1).  
FT VARSPLIC 2501 2535 Missing (in isoform 1).  
FT VARIANT 788 788 /FTid=VSP_007301.  
FT VARIANT 1123 1123 N -> S (in dbSNP:438363).  
FT VARIANT 2016 2016 R -> H (in dbSNP:310632).  
FT VARIANT 2049 2049 L -> P (in dbSNP:3810485).  
FT VARIANT 2049 2049 E -> Q (in dbSNP:3810483).  
FT VARIANT 2049 2049 /FTid=VAR_015600.  
SQ SEQUENCE 2649 AA; 294711 MW; 841892F5521FF459 CRC64;  
  
Query Match 4.2%; Score 98.5; DB 1; Length 2649;  
Best Local Similarity 20.3%; Pred. No. 25;  
Matches 70; Conservative 34; Mismatches 100; Indels 141; Gaps 16;  
  
QY 138 PGLDLEWDFHFTKALVVS-----EEEARLLDRDIEIPVFKTQNKLA 180  
Db 321 PALALE-----FNSSVASGIPSTNYRQRMHQFLYEBA-----AQQLVA 362  
QY 181 DLPSGDIVLKYIYPRKSIATGTPKRLMFDALKAADKFGKVATPLAILLEFIAERAP 240  
Db 363 KLTLRGQVFLKT-----ALQTPALNMLFAPGAL--YAEVPVPSSIMFD---TDQG 408  
QY 241 TLHGHLSCDLVKP---SESRIKVCYMEROLDASIEGIWTLNGRRNDPETLDGLDALRE 297  
Db 409 FLIGRAVSTALVAPVAPDNTVEVLRR--ASSE-----QA 444  
QY 298 LWOLLPVTEGLCPKNCYBPQTSQEQLPFIINFTLSPKS-----ALPEQIYYP 348  
Db 445 LWLLLP-----RCLALGLQPEARLVLEQFQIDPMTFLMHOAVDTLPEQLVVP 496  
QY 349 AF-----GQNDXTIAGLATPFESRGWG-----GLAKS 376  
Db 497 DLPTCALPRWSPVPLRGRNKGELAVLIA-----GWGPGDGRVPPPLIIVGFTGT 551  
QY 377 YPADLASY-----YPDVDLOTANHLOAWISFYSGKPKYMSVLYHT 417  
Db 552 YTLAMASLEVIRPETKVLKICTHNSAADI-----YIREYFHS 589  
  
RESULT 6  
POL2 TRSVR  
ID POL2 TRSVR STANDARD; PRT; 1882 AA.  
AC P25247;
```

```
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DB RNA2 polyprotein (207 kDa protein) [Contains: Coat protein].  
OS Tomato ringspot virus (isolate raspberry) (TomRSV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;  
OC Nepovirus.  
OX NCBI_TaxID=12281;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91311402; PubMed=1856689;  
RA Rott M.E., Tremaine J.H., Rochon D.M.;  
RT "Nucleotide sequence of tomato ringspot virus RNA-2.";  
RL J. Gen. Virol. 72:1505-1514(1991).  
CC -!- SIMILARITY: IDENTICAL FOR THE FIRST 132 AA, AND 75.3% IDENTICAL  
CC FOR THE NEXT 145 AA TO THE RNA1 POLYPROTEIN.  
CC -!- SIMILARITY: TO THE RNA2 POLYPROTEIN OF OTHER NEPOVIRUSES.  
CC -!- CAUTION: It is uncertain whether Met-1 or Met-122 is the  
CC initiator.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC  
CC EMBL; D12477; BAA02043.1; -  
CC F1R; JQ1093; GNVVTR.  
DR InterPro; IPR005054; Nepo_coat.  
DR InterPro; IPR005305; Nepo_coat_C.  
DR InterPro; IPR005306; Nepo_coat_N.  
DR InterPro; IPR008975; Viral_cap_coat.  
DR Pfam; PF03391; Nepo_coat; 1.  
DR Pfam; PF03688; Nepo_coat_C; 1.  
DR Pfam; PF03689; Nepo_coat_N; 1.  
KW Polyprotein; Coat protein; Repeat.  
FT CHAIN 1321 1882 COAT PROTEIN (POTENTIAL).  
FT DOMAIN 554 698 2.5 X TANDEM REPEATS, PRO-RICH.  
FT REPEAT 554 606 1.  
FT REPEAT 607 659 2.  
FT REPEAT 660 698 3 (INCOMPLETE AND APPROXIMATE).  
SQ SEQUENCE 1882 AA; 206802 MW; 0F8958B63AE8DD9D CRC64;  
  
Query Match 4.1%; Score 95.5; DB 1; Length 1882;  
Best Local Similarity 20.3%; Pred. No. 27;  
Matches 84; Conservative 50; Mismatches 129; Indels 151; Gaps 24;  
  
QY 33 WYSTAPMATWAGYDVHAQYKFLC---IHREVIIP-----ALGPVPE 74  
Db 608 WSSPLPLFANFKVNRG-----ACFLQVLPORVVLDFECMDLLSLFQDLPEGPLPS 658  
QY 75 KGQPMHMKSHLTRFGLPFELSFNYSK-----ACFLQVLPORVVLDFECMDLLSLFQDLPEGPLPS 658  
Db 659 -----PSWSS-----PLPLFASFKNRGACFLQVLPARKVSDFMVDVLPFLFSP--VSH 707  
QY 116 TKDDPNTQATRPVLQDL---KAMVPGLDLE-WDFHTKALVYSEEEARLLDRDIEIP 170  
Db 708 QEEPEMVPVLEAADSVDGITEAFDDLECEFSYDSYS-----DEEEAEW-----AEVP 757  
QY 171 VFKTQNKLAADLEPSGDIVLKYIYPRKSIATGTPKRLMFDALKAADKFGKVATPLAIL 230  
Db 758 RCKTMSELCASTLAGD-----AEGLRKSHGVF-----LKLRLVYLIQS 795  
QY 231 LEE-FIAERAPTLHGHLSCDLVKP-----SESRIKVCYMEROLD-----LASTEGIWT 279  
Db 796 FEEPLYSSRA-----FYSVK-VKPVYRPFKEGHIDCTCLDGNNGEWESVDAMWRC 848  
QY 280 NGR-----RNDPETLDGL-----DALRELWOLLPVTE-----GLCPUPN 313  
Db 849 PGLRLNKTFTTRDDWQVQLRIGFNEGRIYRNRVNLNLEEMDLSLHEYFEISSAPVQS 908
```


RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
other Chlamydia strains based on whole genome sequence analysis."
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-histidine + tRNA(His) = AMP +
diphosphate + L-histidyl-tRNA(His).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AEC01648; AAD18802.1; -;
DR EMBL; AEC02172; AAF37970.1; -;
DR EMBL; AB017159; AAF98618.1; -;
DR PIR; H72052; H72052.
DR HSSP; O32422; LQEO.
DR TIGR; CF0084; -;
DR HAMAP; MF_00127; -; 1.
DR InterPro; IPR004154; HGTP anticodon.
DR InterPro; IPR004516; HisS.
DR InterPro; IPR002314; tRNA-synt 2b.
DR InterPro; IPR006195; tRNA ligase II.
DR Pfam; PF03129; HGTP-anticodon; 1.
DR Pfam; PF00587; tRNA-synt 2b; 1.
DR TIGRFAMs; TIGR00442; hisS; 1.
DR PROSITE; PS00862; AA tRNA LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
KW
SQ SEQUENCE 430 AA; 49550 MW; 1892BB7E7C9B74 CRC64;

Query Match 4.1%; Score 94.5; DB 1; Length 430;
Best Local Similarity 22.7%; Pred. No. 4.2;
Matches 64; Conservative 41; Mismatches 126; Indels 51; Gaps 15;

QY 25 SNKHDKWYSTAPMEATMAGAG-YDVHAYKFLCI-----HREVIIPALGPYPEKG 76
Db 104 SHRSNDKFFY-IILMFYERQAGRYQHGFVEAIGVHPLRDAEVLALLWDFSVRG 162

QY 77 QPMHWKSHLFRGLFELSFNYSKSLRFAPEPLGSLTGKDPFNTQAIR-----PVL 130
Db 163 L-QHMOIQNLFLG-GSETRFYDKVLRAVLEKSMGELSALSQQRFSTNVLRLDSKEPED 220

QY 131 QDLKAMVPGU-----DLEWPHFTKALVVSSEEARLLDRDIEIPVKTQNKLAADLE 183
Db 221 QEIIRAPPILDYVSEDEKLYNEILDAL-----RVLEIP-VAINPLRVRLGLD 267

QY 184 PSGDIVLK-TYIYPRIKSIATGPKERLMPDAIKAADKFGKATPLAILEEFTIAERAPTL 242
Db 268 YSDDLVEATTTQEVSYALGGGR---YDGLISA--FGASLPACGFGVGLERAIQTL 321

QY 243 LGHFLSCLLVKPS-ESRIKYCHERQLDLASIEGIWTLNGRR 283
Db 322 LAQ-----KRIEPOFFPHKRLIPMEPDADQFCLS--WSOHLRR 357

RESULT 9
POL-HTL1A STANDARD; PRT; 896 AA.
AC P03362;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE POL polyprotein [Contains: Reverse transcriptase (EC 2.7.7.49);
DE Ribonuclease H (EC 3.1.26.4)].
GN POL.
OS Human T-cell leukemia virus type I (strain ATK) (HTLV-I).
OS Viruses; Retroviridae; Deltaretrovirus.

OX NCBI_TaxID=11926;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83221647; PubMed=6304725;
RA Seiki M., Hattori S., Hirayama Y., Yoshida M.;
RT "Human adult T-cell leukemia virus: complete nucleotide sequence of
the provirus genome integrated in leukemia cell DNA";
Proc. Natl. Acad. Sci. U.S.A. 80:3618-3622(1983).
RL [2]
RN SEQUENCE OF 69-185 FROM N.A.
RX MEDLINE=89210803; PubMed=2468487;
RA Bangham C.R.M., Daenke S., Phillips R.E., Cruickshank J.K.,
RA Bell J.I.;
RT "Enzymatic amplification of exogenous and endogenous retroviral
sequences from DNA of patients with tropical spastic paraparesis";
EMBO J. 7:4179-4184(1988).
RL -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
phosphomononucleotide.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ {DNA} (N).
CC -!- PTM: Cleavage sites that yield the mature proteins remain to be
determined.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; J02029; AAA96673.1; -;
DR EMBL; X14144; CAA32360.1; -;
DR PIR; A03961; GNLJGH.
DR HSSP; P03355; 1MWL.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVise.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF00075; rnaaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
DR Polyprotein: Hydrolase; Endonuclease; Nuclease; Transferase;
KW RNA-directed DNA polymerase.
KW
SQ SEQUENCE 896 AA; 100141 MW; 113D45D4BD79C65F CRC64;

Query Match 4.1%; Score 94.5; DB 1; Length 896;
Best Local Similarity 20.7%; Pred. No. 12;
Matches 79; Conservative 62; Mismatches 151; Indels 89; Gaps 18;

QY 71 PYPEKGQPMHWKSHLFRGLFELSFNYSKSLRFAPEPLGSLTGKDDP--ENTQAIRP 128
Db 123 PLPKQFQY-----FAFTVPOOCNYGPG-TRYAMKVLPO--GFRNSPTLFEMQ-LAH 170

QY 129 VLQDLKAMVPGU-LWEPDHFHTKA-----LVVSEEEARTLLDRDIEIPVKTQNKLA 180
Db 171 ILQPIRQAPFQCTILQYVDDILLASPSHEDLLILSEATMASLISHGLVSENKIQTPGT 230

QY 181 DLEPSGDIVLKYIYPRIKSIATGPKERLMPDAIKAADKFGKATP--LAILEE--PIA 236
Db 231 -----IKFLQIISPNHLYDAVPTPIRSWALPELQALLGIQWVS 273

QY 237 ERAPTLTG--HFLSCLLVKPSERIKYCHERQL--DLASIEGIWTLNGRRNDPETLGLD 293
Db 274 KGTPTLRQLRSLYCALORHDPDRQIYLNPSQVSLVQLRQALSONCESRLVQLTPLL 333

QY 294 ALRELWQLLPVTEGLCPLENCFYEPGTS-----PQEQLPFI-INFTLSPKALPEPQIYF 347
Db 334 AI-----MLTLT-----GTTTVVFQSQEQWPLVWLHAPLPHPTSQCPWQGLLA 375

QY 348 PAFGQNDKTIABGLATFFESRCGWGLAKSYPADLASVYPDVDTANHLQAWISFYK 407

```
Db 376 SAVLLDK-----YTLQSYGLLQOTIHNISTQTENQFIQTSHPVS----- 417
QY 408 KPYNSVYLHTEAFSAQAQEV 428
Db 418 -PILHSHRPNKLGATGEL 437

RESULT 10
SVV_SVNY3
ID_SV SYV SYNY3 STANDARD; PRT; 910 AA.
AC Q55322;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
GN VALS OR SLR0557.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
CC + L-valyl-tRNA(Val).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D64006; BAA10881.1; -
CC PIR; S76034; S76034.
CC HSP; P96142; IGAX.
CC InterPro; IPR002300; tRNA-synt_la.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR002303; tRNA-synt_val.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00986; tRNA-synt_1.
CC TIGR; TIGR00422; vals; 1.
CC PROSITE; PS00178; AA_tRNA_Ligase_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 46 56 "HIGH" REGION.
FT SITE 539 543 "KMSKS" REGION.
FT BINDING 542 542 ATP (BY SIMILARITY).
FT BINDING 910 AA; 102737 MW; 4909BCD889BF42E CRC64;
SQ SEQUENCE
Query Match 4.0%; Score 94; DB 1; Length 910;
Best Local Similarity 22.1%; Pred. No. 13;
Matches 83; Conservative 46; Mismatches 117; Indels 130; Gaps 22;
QY 63 EVIIPALGPYKPGQPMHMKSHLT---RGLPFELSFNYSKSLRFAPEPGLSLTGTKDD 119
Db 262 ELVDPEFGTCVKVTPAHDPNDPVMGQRHNLFPINLNRDGLS-----NENGDFAG--QD 315
QY 120 PFTQAIKRVPLDLKAMVPLGLDLEWFDHFTKALVSYSEEARLL---DRDIEIPVFKTON 176
Db 316 RF--EARKNVQALEA-----QGLVKIPYRHSVPYGDGK-KVPV----- 353
QY 177 KLAADLEPSGDIVLXTIYIPRIKSIATGTPKRLMFDIAKADKPKVATPLALEZ--- 233
```

```
Db 354 -----EP-----LUSTQWFKIESLAQNA-----LACDEDNS 381
QY 234 --FIAERAPTLGHFLSCDLVKPSESRIKVCYMERQDLASIEGIWTNGRNDPFDLDG 291
Db 382 PNFVPERGKGYRDLW-----VKLKWICISQL-----WM--GHQ----- 414
QY 292 LDALRELWQLLPVTEGCLPNCPEYPGTSPQBLPFIINFTLSPKALPFPQIYFFAPG 351
Db 415 ----IPAWYVISETNG-----AITDTPFIVAYDEAEALAKAKAE-YGPTVQ 456
QY 352 -QNDKTIAEGLATPFESRGMGLAKSYPA---DLASYVPDVLQTA-NHQAWI----- 400
Db 457 LQDDPDV---LDTWFSGLWFFSTMGWPEQTDLLAKYPTSTLVTFDIIFFVAVRMTMM 513
QY 401 SFSYKGGKPYMSVYLH 416
Db 514 AGHFTGQIPFKDVIYH 529

RESULT 11
SYL_STRPN
ID_SYL_STRPN STANDARD; PRT; 833 AA.
AC Q97SS0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR SP0234.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetelin H., Nelson K.B., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen, C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE007338; AAK74433.1; -
CC PIR; H95029; H95029.
CC TIGR; SP0254; -
CC HAMAP; MF_00049; -; 1.
CC InterPro; IPR002302; Leu-tRNA-syntla.
CC InterPro; IPR002300; tRNA-synt_la.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR009008; ValRS_1IleRS_edit.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; tRNA-synt_1.
CC TIGR; TIGR00396; leus_bact; 1.
CC PROSITE; PS00178; AA_tRNA_Ligase_I; 1.
```

KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 41 52 "HIGH" REGION.
 FT SITE 610 614 "KMSKS" REGION.
 FT BINDING 613 613 ATP (BY SIMILARITY).
 SQ SEQUENCE 833 AA; 94421 MW; 7CAA4FB5B4F99B83 CRC64;

Query Match 4.0%; Score 93; DB 1; Length 833;
 Best Local Similarity 19.1%; Pred. No. 14; Indels 148; Gaps 26;
 Matches 95; Conservative 66; Mismatches 188; Indels 148; Gaps 26;

QY 16 YVLSQALN-SNKHDKTW--WYSTAMPATWAGAGYDVHAYKFL-----CHREVIIP- 67
 Db 119 YDMDREVNTDNYKWTQWIFT-----KLYEKLAYEAEPVNVWBEELGTALANEVLPD 174
 QY 68 ---ALGPYKEGQPM-HWKSHTLRFQ--LPPFL-SFNYSKSLRLRAFEPLGSLTGTK--- 117
 Db 175 GTSERGYPVVRKPMQKMLKITAYAEHLNDLDELWSESNDQMRNWIGKSTGANVTF 234
 QY 118 ----DDPENTQAIRPVL---QDLKAMVPGDLDFHFTKALVSEEEARTLLDRDIEI 169
 Db 235 KVKGTCKEFTVTRPTDLPFGATFTVLAP-----EHELVDAITSEQAQAVD----- 282
 QY 170 PVFKTONKLAADLEPGSDIVLKYIYPRKSIATGTPKRLMPDAKAAADKFGKATPLA 229
 Db 283 --YHQASLKSGLDARTDLAKETGVTGAYAINPVNGKEMPIWIADYVLASYGTGAV--- 337
 QY 230 ILBEFTAERAAPTLLGHFLSCDLVKPSESRIKVCMERQLDLASI-----EGITWLN 280
 Db 338 -----MAVPAHQ--RDWEFAKQDLPIVEVLEGGNVEEAAYTED 375
 QY 281 GRNDPPTDGL--DALREL--WQLPVTEGLC-----PLPN 313
 Db 376 GLHVNDFDLGNKEDAIKIVAW----LEEKGGQEKVTYRLRWLFSRQRYWGPPIPI 431
 QY 314 CFVEPGTS---POEQLPFIINFLSPKSPALPEQIYFPAQGND---KTAIEGLATFPES 367
 Db 432 IHWDGTSVAVPETELPVLVPT---KDIRPSTGSGPLANLTDWLEVTREDGVKGRRT 488
 QY 368 R---GMGGLAKSY-----PADLASYPVDVLDQTAN-----HL---QAWISF 402
 Db 489 NTMEQWAGSSWYLYRVIDPHNTEKLADEDLKQWLPVDIYVGGAHVLHLLVAREWHKF 548
 QY 403 SY-----KGKPKYSVY 414
 Db 549 LYDLGVVPTKEPQKLF 565

RESULT 12

POLN_EBVT
 ID POLN_EBVT STANDARD; PRT; 2492 AA.
 AC P27282;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nonstructural polyprotein [Contains: Nonstructural protein NSP1;
 DE Nonstructural protein NSP2; Nonstructural protein NSP3; Nonstructural
 DE protein NSP4].
 DE Venezuelan equine encephalitis virus (strain Trinidad donkey).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 CX NCBI_TaxID=11038;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89243175; PubMed=2524126;
 RA Kinney R.M., Johnson B.J.B., Welch J.B., Tsuchiya K.R., Trent D.W.;
 RT "The full-length nucleotide sequences of the virulent Trinidad donkey
 RT strain of Venezuelan equine encephalitis virus and its attenuated
 RT vaccine derivative, strain TC-83,"
 RL Virology 170:19-30 (1989).
 CC -!- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS

CC BETWEEN THE CODONS FOR 1879-GLN AND 1880-ARG.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; J04332; AB02518.1; --
 DR PIR; A31467; MNWVTD.
 DR MEROPS; C09.001; --
 DR InterPro; IPR002589; Alpp.
 DR InterPro; IPR002620; Peptidase C9.
 DR InterPro; IPR001788; RNA dep RNAPol2.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PS_vir.
 DR InterPro; IPR000606; Viral_helicase.
 DR Pfam; PF01661; Alpp; 1.
 DR Pfam; PF01707; Peptidase C9; 1.
 DR Pfam; PF00978; RNA dep RNAPol2; 1.
 DR Pfam; PF01443; Viral_helicase; 1.
 DR SMART; SM00506; Alpp; 1.
 DR Polyprotein; Nonstructural protein; RNA-binding; ATP-binding;
 KW Helicase.
 FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
 FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
 FT CHAIN 1330 1879 NONSTRUCTURAL PROTEIN NSP3.
 FT CHAIN 1880 2492 NONSTRUCTURAL PROTEIN NSP4.
 FT NP_BIND 721 728 ATP (POTENTIAL).
 SQ SEQUENCE 2492 AA; 277902 MW; 1BAD415B70DC3FA0 CRC64;

Query Match 4.0%; Score 93; DB 1; Length 2492;
 Best Local Similarity 20.9%; Pred. No. 64;
 Matches 67; Conservative 45; Mismatches 91; Indels 118; Gaps 18;

QY 57 FLC-IHREVIIPALGPYKEGQPMHWKSHLRFGLP-----FELSFNYSKSLRLRAFPAP- 109
 Db 2206 YLGGHRLV-----RRLNAVLLPNHITLFDMSAEDFDIAIAHFQPG 2248
 QY 110 -----LGSUTGKDDPNTQAIRPVLDLQKAMVPGDLDFHFTKALVSEEEARTL 162
 Db 2249 DCVLETDIASFDKSEDDAMALTALM-ILEDL-----GVDAELL-----TL 2287
 QY 163 LD-----RDIEIPVFKTONKLAADLEPSG-----DIVLKYIYPR- 198
 Db 2288 IEAFAFGESSIHLPT-KTKFKGAMMK-SGMFLTLFVNTVINIVIASVLRERLTGSPCA 2345
 QY 199 -----KSIATGTPKRLMFDKAKDKFGKQVAT-----PLAILEBFIAERAPTLGHFLSC 249
 Db 2346 AFIGDNDIVKGVSKDLNAD-----RCATWLNMEVKIIDA VVGKAPVFCGFI LC 2396
 QY 250 DLVKPSEERI-----KYCYMERQLDLASIEGIWTLNGRNDPDTLGLDALRE---LWQL 301
 Db 2397 DSVTGTACRVADPLKRLFKLQKPL-----AADDEHDDRRRALHEESTRWNR 2443
 QY 302 LPVTEGLCPNPCEYEP-GTS 321
 Db 2444 VGILSELCKAVESRYETVGTGS 2464

RESULT 13

ADHE_CLOAB
 ID ADHE_CLOAB STANDARD; PRT; 862 AA.
 AC P33744; Q45808; Q45809;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Aldehyde-alcohol dehydrogenase [includes: Alcohol dehydrogenase
 DE (EC 1.1.1.1) (ADH); Acetaldehyde dehydrogenase [acetylating]
 DE (EC 1.2.1.10) (ACDH)].
 GN ADHE OR AAD OR CAP0162.

OS Clostridium acetobutylicum.
 OG Plasmid pSOL1.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_taxid=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=94042861; PubMed=8226639;
 RA Fischer R.J., Helms J., Duerre P.;
 RT "Cloning, sequencing, and molecular analysis of the sol operon of
 RT Clostridium acetobutylicum, a chromosomal locus involved in
 RT solventogenesis.";
 RL J. Bacteriol. 175:6959-6969(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=94131967; PubMed=8300540;
 RA Nair R.V., Bennett G.N., Papoutakis E.T.;
 RT "Molecular characterization of an aldehyde/alcohol dehydrogenase gene
 RT from Clostridium acetobutylicum ATCC 824.";
 RL J. Bacteriol. 176:871-885(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Onelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -!- FUNCTION: THIS ENZYME HAS PROBABLY TWO ACTIVITIES: ADH, AND ACDH.
 CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
 CC NADH.
 CC -!- CATALYTIC ACTIVITY: Acetaldehyde + CoA + NAD(+) = acetyl-CoA +
 CC NADH.
 CC -!- INDUCTION: Induced during switch to solvent production.
 CC -!- SIMILARITY: TO THE C.KLUVERI NADP-DEPENDENT SUCCINATE
 CC SEMIALDEHYDE DEHYDROGENASE (SUCD) (IN THE N-TERMINAL SECTION).
 CC -!- SIMILARITY: In the C-terminal section; belongs to the iron-
 CC containing alcohol dehydrogenase family.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; X72831; CAA51344.1; -;
 DR ENBL; L44817; AAD04638.1; -;
 DR ENBL; A8001438; AAK76907.1; -;
 DR PIR; A49346; A49346.
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR PIR; IPR001670; Fe-ADH.
 DR Pfam; PF00171; aldedh; 1.
 DR Pfam; PF00465; Fe-ADH; 1.
 DR PROSITE; PS00913; ADH_IRON_1; 1.
 DR PROSITE; PS00060; ADH_IRON_2; FALSE_NEG.
 DR Oxidoreductase; Multifunctional enzyme; NAD; Plasmid;
 KW Complete proteome.
 FT ACT_SITE 244 244 BY SIMILARITY.
 FT NP_BIND 420 425 NAD (POTENTIAL).
 SQ SEQUENCE 862 AA; 95321 MW; E09E32B8DD08B0 CRC64;
 Query Match 4.0%; Score 92.5; DB 1; Length 862;
 Best Local Similarity 20.4%; Pred. No. 16;
 Matches 78; Conservative 48; Mismatches 127; Indels 129; Gaps 18;

Db 444 ERENMLW-----FRPHKVIKFP--CCLQFALKDLKDLKKKRAFIVTDSDPYNLNVYD 495
 QY 128 PVLQDLKAMVFGDLDFHFKALVSEEEARTLLDRDIEIPVFKTONKLAADLEP--- 184
 Db 496 SIIKILE-----HLDID-EKVENK--VGREADLKTIKATEEMSSFMPTIIALGCTPEMS 548
 QY 185 SGDIVLKYIYIPRI-----KSIATGTPK--ERLMPDIALKAADFQGVATP-- 227
 Db 549 SAKLMMVLYEHPKVFEDLAIFKMDIRKRIYT-FPKLGKKAMLVATTSAGSGSEVTPFA 607
 QY 228 -----LAILESFIADRAPTLHGHLFSLCD-LVKPSESRIKVVYVME 265
 Db 608 LVTDNNTGKMYLADYEMTPNNAIVDAELMMKMPKGLTAYSGIDALVNSIEAYTSVYASE 667
 QY 266 RQLDLASIEGIWTL-----NGRRND-----PETLDGLDALRELWQLL----- 302
 Db 668 YTNGLA-LEAIRLIFKYLPEAYKNGRTNEKAREKVAHAHASTMAGNASANAFGLCHSMAIK 726
 QY 303 -----PVTEGLCP--LPNCFYE-----PGTS 321
 Db 727 LSSEHNTPSGIANALLIEEVKFNVDNPKVQAPCPQYKYPNTIFRYARIADYIKLGNT 786
 QY 322 PQQLPFTINFTLSPKSGALPEP 343
 Db 787 DEEKVYDLLINKIHELKALNIP 808
 RESULT 14
 YMA8_CAEEL STANDARD; PRT; 282 AA.
 ID YMA8_CAEEL
 AC P34453;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein F54P2.8 in chromosome III.
 GN F54P2.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lighthouse J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 CC -!- SUBCELLULAR LOCATION: Peroxisomal (Potential).
 CC -!- SIMILARITY: Belongs to the PXP / PEX19 family.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; L23645; AAK36133.1; -;
 DR PIR; S44825; S44825.0000

OV 74 FKGPDMFWKSHI TRCFI DEFI SNYSKSI I DEAFEDICSI NT

DR InterPro: IPR006708; Pex19.
DR Pfam: PF04614; Pex19; 1.
KW Hypothetical protein; Peroxisome; Prenylation; Lipoprotein.
FT DOMAIN 270 275 Poly-ALA.
FT LIPID 279 279 S-farnesyl cysteine (Potential).
SQ SEQUENCE 282 AA; 30988 MW; 681FBDFF8F08DBE4 CRC64;

Query Match 4.0%; Score 92; DB 1; Length 282;
Best Local Similarity 23.1%; Pred. No. 3.7; Mismatches 104; Gaps 16;
Matches 67; Conservative 37;

QY 154 VSEEARITLDRDIEIPVKTQNKLAADLEPS-GDIVLKYIYPIRIKSIAGTTPKERLMP 212
DB 1 MTDETTQNIKD-----KTEELAALDQTLGEFT-----ATPAPKPRTTD 40

QY 213 D-----ATKADKGGKVATPLAILEETIARRAPLLGHFLSCDLVKPSESRI 259
DB 41 DELDELMASDAQEAAKAKDKQKMLEQNVTLQEEAMKAG-----ADPSEGE- 88

QY 260 KVCYMERQDLASIEGIWTLNGRRNDPDTLDGLDALRELWQL-----PVTEGLC 309
DB 89 -----GQQLD-----PNDPEALAMMDALKQLMCESSNVANASNEPEFMAGLD 131

QY 310 PL--PNCFYBPGTSPOEQLPFFINTLSPKSL--PEFOIY--FPAP-----GQNDKT- 356
DB 132 MLRSPN-----SPMEPTFMSIMQTLASKEVYVPLKEIFDNYPKYLEDNAGLDATK 184

QY 357 -----IAEGLATPFESRGGWGLAKSYPADLASI-YPDVLDQTANHLQ 397
DB 185 ERYEKQFEVLGKICTEFEXQ--PELAEVQPVDAATQPAEADPASTIEHFE 232

DR HSP; P32168; IE94.
DR HAMAP; MF 00249; -. 1.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003593; AAA_ATPase_cent.
DR InterPro: IPR001270; Chaperonin clpA/B.
DR InterPro: IPR004491; Hsp_HslVU.
DR Pfam; PF00004; AAA; 1.
DR PRINTS; SM00300; CLPPTOTASEA.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00390; hslU; 1.
KW Chaperone; ATP-binding; Complete proteome.
FT NP_BIND 59 66 ATP (POTENTIAL).
SQ SEQUENCE 443 AA; 50164 MW; 6AF0F90849B686DA CRC64;

Query Match 4.0%; Score 92; DB 1; Length 443;
Best Local Similarity 24.6%; Pred. No. 7;
Matches 73; Conservative 44; Mismatches 116; Indels 64; Gaps 17;

QY 119 DPFTQAIRPVLDLKAWP-GLDLEWFDHFTKALVSEEE-ARTLLD-RDRIIPVFKTQ 175
DB 123 DKIEEAVVEVETAKLLPFLPSGVSEKKQBYANSLKQOQRIAGGELDSREIEIEVRKS 182

QY 176 NKLAADLEPSGDIVLKTYIYPIRIKSIATGTPKERLMFDAIKAADKFGKVATPLAILEEFTI 235
DB 183 IEIDSNVYPP-----ILRV-----QENLIKVFHKEQDKVKKTLKSVKEAKEALK 225

QY 236 AERAPTL-CHFLSCDLVKPSESRIKVCYMERQDLASIEGIWTLNGRRNDPDTLDGLDA 294
DB 226 AEISDTLLDGEAIKVGKLGKRAESSGVIFIDE--IDKIAVS---SKEGSDQD-SKEGVQR 279

QY 295 LRELWQLLVTEGLCPNCFYBPGTSPOEQLPFI--NFTLS-PKSALPEPQIYFP--- 348
DB 280 -----DLLPIVEG--SVNTKY--GSIKTEHILFIAAGAFHLSKPSDLPELQGRFFLRV 330

QY 349 -----AFQNDKTI-----AEGLATPFESRGGWGLAK-SYPADLAS 383
DB 331 ELENLTBEIMVMILTQTKTSIIKQYQALLKVGVEIAFEDDAIKELAKLSYNANQKS 387

Search completed: July 22, 2004, 18:45:16
Job time : 22 secs

DR InterPro: IPR006708; Pex19.
DR Pfam: PF04614; Pex19; 1.
KW Hypothetical protein; Peroxisome; Prenylation; Lipoprotein.
FT DOMAIN 270 275 Poly-ALA.
FT LIPID 279 279 S-farnesyl cysteine (Potential).
SQ SEQUENCE 282 AA; 30988 MW; 681FBDFF8F08DBE4 CRC64;

Query Match 4.0%; Score 92; DB 1; Length 282;
Best Local Similarity 23.1%; Pred. No. 3.7; Mismatches 104; Gaps 16;
Matches 67; Conservative 37;

QY 154 VSEEARITLDRDIEIPVKTQNKLAADLEPS-GDIVLKYIYPIRIKSIAGTTPKERLMP 212
DB 1 MTDETTQNIKD-----KTEELAALDQTLGEFT-----ATPAPKPRTTD 40

QY 213 D-----ATKADKGGKVATPLAILEETIARRAPLLGHFLSCDLVKPSESRI 259
DB 41 DELDELMASDAQEAAKAKDKQKMLEQNVTLQEEAMKAG-----ADPSEGE- 88

QY 260 KVCYMERQDLASIEGIWTLNGRRNDPDTLDGLDALRELWQL-----PVTEGLC 309
DB 89 -----GQQLD-----PNDPEALAMMDALKQLMCESSNVANASNEPEFMAGLD 131

QY 310 PL--PNCFYBPGTSPOEQLPFFINTLSPKSL--PEFOIY--FPAP-----GQNDKT- 356
DB 132 MLRSPN-----SPMEPTFMSIMQTLASKEVYVPLKEIFDNYPKYLEDNAGLDATK 184

QY 357 -----IAEGLATPFESRGGWGLAKSYPADLASI-YPDVLDQTANHLQ 397
DB 185 ERYEKQFEVLGKICTEFEXQ--PELAEVQPVDAATQPAEADPASTIEHFE 232

DR HSP; P32168; IE94.
DR HAMAP; MF 00249; -. 1.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003593; AAA_ATPase_cent.
DR InterPro: IPR001270; Chaperonin clpA/B.
DR InterPro: IPR004491; Hsp_HslVU.
DR Pfam; PF00004; AAA; 1.
DR PRINTS; SM00300; CLPPTOTASEA.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00390; hslU; 1.
KW Chaperone; ATP-binding; Complete proteome.
FT NP_BIND 59 66 ATP (POTENTIAL).
SQ SEQUENCE 443 AA; 50164 MW; 6AF0F90849B686DA CRC64;

Query Match 4.0%; Score 92; DB 1; Length 443;
Best Local Similarity 24.6%; Pred. No. 7;
Matches 73; Conservative 44; Mismatches 116; Indels 64; Gaps 17;

QY 119 DPFTQAIRPVLDLKAWP-GLDLEWFDHFTKALVSEEE-ARTLLD-RDRIIPVFKTQ 175
DB 123 DKIEEAVVEVETAKLLPFLPSGVSEKKQBYANSLKQOQRIAGGELDSREIEIEVRKS 182

QY 176 NKLAADLEPSGDIVLKTYIYPIRIKSIATGTPKERLMFDAIKAADKFGKVATPLAILEEFTI 235
DB 183 IEIDSNVYPP-----ILRV-----QENLIKVFHKEQDKVKKTLKSVKEAKEALK 225

QY 236 AERAPTL-CHFLSCDLVKPSESRIKVCYMERQDLASIEGIWTLNGRRNDPDTLDGLDA 294
DB 226 AEISDTLLDGEAIKVGKLGKRAESSGVIFIDE--IDKIAVS---SKEGSDQD-SKEGVQR 279

QY 295 LRELWQLLVTEGLCPNCFYBPGTSPOEQLPFI--NFTLS-PKSALPEPQIYFP--- 348
DB 280 -----DLLPIVEG--SVNTKY--GSIKTEHILFIAAGAFHLSKPSDLPELQGRFFLRV 330

QY 349 -----AFQNDKTI-----AEGLATPFESRGGWGLAK-SYPADLAS 383
DB 331 ELENLTBEIMVMILTQTKTSIIKQYQALLKVGVEIAFEDDAIKELAKLSYNANQKS 387

Search completed: July 22, 2004, 18:45:16
Job time : 22 secs

RESULT 15
HSLU HELPJ
ID HSLU HELPJ STANDARD; PRT; 443 AA.
AC Q9ZLW1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent hsl protease ATP-binding subunit hslU.
GN HSLU OR JHP0465.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: Chaperone subunit of a proteasome-like degradation
CC complex (By similarity).
CC -!- SUBUNIT: A double ring-shaped homohexameric of hslU is capped on
CC each side by a ring-shaped hslU homohexameric (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the clpx chaperone family. HslU subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001480; AAD06034.1; -.
DR PIR; F71929.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 18:41:02 ; Search time 39 Seconds
(without alignments)
3535.422 Million cell updates/sec

Title: US-10-099-704-2
Perfect score: 2322
Sequence: 1 MEISKKAATLIPKPVLSQ.....PEAFSAQAQEVAMCHDGHNP 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

ALIGNMENTS

RESULT 1

O94204 PRELIMINARY; PRT; 448 AA.
AC O94204;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE Dimethyl-allyl-tryptophan-synthase.
GN D1.
OS Claviceps purpurea (Ergot fungus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
OX NCBI_TaxID=5111;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P1;
RX MEDLINE=99168777; PubMed=10071219;
RA Tudzynski P., Hoelter K., Correia T.H., Arntz C., Grammel N.,
RA Keller U.;
RT "Evidence for an ergot alkaloid gene cluster in Claviceps purpurea.";
RL Mol. Gen. Genet. 261:133-141(1999).
DR EMBL; AJ011963; CAB39314.1; -.
SQ SEQUENCE 448 AA; 51672 MW; 625709FAAF269242 CRC64;

Query Match 39.3%; Score 911.5; DB 3; Length 448;
Best Local Similarity 43.7%; Pred. No. 9.5e-69;
Matches 181; Conservative 71; Mismatches 151; Indels 11; Gaps 4;
QY 15 FYVLSQALNLGNKDHTRKWWYSTAPMFATMAGAGYDVHAQYKFLGCIHREVIIPALGPYPE 74
Db 12 YEILSLIFDFSNQELWHSHTAPMFAAMLDNAGYNIHQVRLHGFKKHIIIPFLGVYPT 71
QY 75 KQCPMHWKSHLFRGLPFELSPNYSKSLRFAFPLGSLTGTGKDPFNTQAIRPVLQDLK 134
Db 72 KDKK-RWLSILTRGLPLELNLCTDSVVRVTEPINEVTGTEKDTFTNLAIMTSVQKLA 130
QY 135 AMVPGLDLEWDFHFKALVWSEEEARTLLDRDIEIPVFKTONKLAADLEPSGDIVLKYTI 194
Db 131 QIQAGIDLEWTSYFFKDELTLDESEATLQSNELVKEQIKTONKLAIDLKES-QPALKVVF 189

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	911.5	39.3	448	3	O94204
2	908.5	39.1	448	3	Q9C141
3	895.5	38.6	455	3	Q12594
4	849.5	36.6	441	3	Q9C140
5	202.5	8.7	187	3	Q9C451
6	117	5.0	577	10	Q8S922
7	112	4.8	1772	3	Q96X17
8	110.5	4.8	462	17	Q8IXF1
9	107.5	4.6	490	5	Q21549
10	104	4.5	595	5	Q7YZB8
11	102	4.4	1043	4	Q86W25
12	101	4.3	503	11	Q8XOC4
13	101	4.3	503	11	Q8BSQ7
14	101	4.3	878	10	Q42710
15	101	4.3	1621	16	Q9KTA5
16	100	4.3	323	16	Q9A9D7

```

QY 195 YPRIKSIATGTPKRLMFDAL-KAADKFGKVAATPLAILEEFIAERAPT-----LLGH 245
Db 190 YPHLKSIATGKSTHDLIFDSVLKLSQKSDSIQAFQVLCDYVSRRHSAEVDQGHALHAR 249

QY 246 FLSCDLVKPESRIKVCYMERQDLASIEGIWTLNGRNDPETHDGLDALRELWQLLPVT 305
Db 250 LLSDDLIDPAKSRVKIYLKXTVSLWSMEDLWLGQGVDASTWMDGLDMLRELWSLLKVP 309

QY 306 EGLCPLNCFYEPGTSPOEQLPFIINFTLSPKSLPQIYFPAPQNDKTIAGELATFF 365
Db 310 TGHLEYPKGYLDELGEIPNEQLPSMANYTLHHNDPMPEPQVYFVFGMNDAEISNALTIF 369

QY 366 ESRGWGLAKSYPADLASYPVDVLOTANHLQAWISFSYKGGKPYMSVYLHTE 419
Db 370 QRHGFDDMAKNYRVFLQDSYPYHDFESLNYLHAYISFSYRRNKPYSVYLHTE 423

RESULT 2
Q9C141
ID Q9C141 PRELIMINARY; PRT; 448 AA.
AC Q9C141
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DE Putative dimethyl-allyl-tryptophan-synthase.
GN CPD2.
OS Claviceps purpurea (Ergot fungus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
OX NCBI_TaxID=5111;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T5;
RA Artz C., Tudzynski P.;
RT "Molecular analysis of dimethyl-allyl-tryptophan-synthase-genes.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBD databases.
DR EMBL; AJ312753; CAC37396.1; -.
SQ SEQUENCE 448 AA; 51520 MW; 1ED9FE8900810B59 CRC64;

Query Match 39.1%; Score 908.5; DB 3; Length 448;
Best Local Similarity 44.0%; Pred. No. 1.7e-68;
Matches 182; Conservative 65; Mismatches 156; Indels 11; Gaps 4;

QY 15 FVYLSQALNLSKNDHTKWWYSTAPMTWAGAGYDVHQAQKFLCIHREVIIPALGPYPE 74
Db 12 YEILSLIFDPPSNEQRLWHSAPMAALDNAGYSVHDQYRHLISFIFTHIIPFLGVYPT 71

QY 75 KGQPMHWKSHLTFRGLPFELSFNYSKSLRFAEPGLSLTGTKDPPFNTQATRPVLQDLK 134
Db 72 KGQE-RWLSILTRCGFLFELSNCDSVVRAYEPINEMTGEXDPSNTPLIGSVQKLA 130

QY 135 AMVPGDLDFWDFHTKALVVSSEEARLLDRDIEIPVFQTKNLAADLEPSGDIVLTKTYI 194
Db 131 QIQAGIDLEWFSYFKDELTDSESAILODELVEKQIKTQNKALDLKES-QFALKYVF 189

QY 195 YPRIKSIATGTPKRLMFDAL-KAADKFGKVAATPLAILEEFIAERAPT-----LLGH 245
Db 190 YPHLKSIATGTHLIFDSVFLSKSDSIQAFQALCDYVSRRNDSSEVDQGHALHAR 249

QY 246 FLSCDLVKPESRIKVCYMERQDLASIEGIWTLNGRNDPETHDGLDALRELWQLLPVT 305
Db 250 LLSDDLIDPAKSRVKIYLKXTVSLWSMEDLWLGQGVDASTWMDGLDMLRELWSLLKVP 309

QY 306 EGLCPLNCFYEPGTSPOEQLPFIINFTLSPKSLPQIYFPAPQNDKTIAGELATFF 365
Db 310 TGHLEYPKGYLDELGEIPNEQLPSMANYTLHHNDPMPEPQVYFVFGMNDAEISNALTIF 369

QY 366 ESRGWGLAKSYPADLASYPVDVLOTANHLQAWISFSYKGGKPYMSVYLHTE 419
Db 370 QRHGFADMAKNYRVFLQDSYPYHDFESLNYLHSLVFSYRRNKPYSVYLHTE 423

```

```

RESULT 3
Q12594
ID Q12594 PRELIMINARY; PRT; 455 AA.
AC Q12594
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Dimethylallyl-trans-transferase (EC 2.5.1.1).
GN DMAP.
OS Claviceps fusiformis (Ergot fungus).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae; Claviceps.
OX NCBI_TaxID=40602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96067540; PubMed=7488077;
RA Tsai H.F., Wang H., Gebler J.C., Poulter C.D., Schardl C.L.;
RT "The Claviceps purpurea gene encoding dimethylallyltryptophan
RT synthase, the committed step for ergot alkaloid biosynthesis.";
RL Biochem. Biophys. Res. Commun. 216:119-125(1995).
DR EMBL; L39440; AAC18993.1; -.
DR GO; GO:0004161; F:dimethylallyltransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
SQ SEQUENCE 455 AA; 51858 MW; A913F8B60EBBEB63 CRC64;

Query Match 38.6%; Score 895.5; DB 3; Length 455;
Best Local Similarity 42.1%; Pred. No. 2.2e-67;
Matches 187; Conservative 75; Mismatches 151; Indels 31; Gaps 9;

QY 3 ISKXAATLLPKPFYVLSQLNLSKNDHTKWWYSTAPMTWAGAGYDVHQAQKFLCIHR 62
Db 2 MTKAPATAV---YDTLSLLDFDFNQEQRLWHSIAPMAALDTAGNVHDQYEHGIFK 58

QY 63 EVIIPALGPPEKQPMH-KWSHLTRFGLPFELSFNYSKSLRFAEPGLSLTGTKDDPP 121
Db 59 KHIIFPLGVYPAQKK--HTWPSVLTRYGIPFELSNCDSVVRAYETFEPTTEHTGTGDSY 116

QY 122 NTOAIRPVLQDLKAMPGLDLEWDFHTKALVVSSEEARLLDRD-IEIPVFQTKNLA 180
Db 117 NAFALLEICQLVRIQPGIDMEWFSYFRNELVLNATESARLGNDSVNOQPIPTONKAL 176

QY 181 DLPSGD-IVLKYIYIPRIKSIATGTPKRLMFDAL-KAADKFGKVAATPLAILEEFIAER 238
Db 177 DLK--GDRFALKVLYLPHLKSATGVSSSHDLIENSVRKLSQKHTSIQPSFNVLCDYVASR 234

QY 239 -----AFTLLGHFLSCDLVKPSSRIKVCYMERQDLASIEGIWTLNGRR 283
Db 235 NDPDSNAEABAGVPASALRALLSCDLVDPSKRIKIYLLGQTVSLTAWEDLWTLGGR 294

QY 284 NDPETDGLDALRELWQLLPVTEGLCPFPNCFYEPGTSPOEQLPFIINFTLSPKSLPPE 343
Db 295 TDSSTLNGLDMMRELWELLQIPSGFMKYPESDLKGEVDEQLPSMVHYALHPDQMPPEP 354

QY 344 QIYFPAGQNDKTIAGELATFFESRCGGGLAKSYPADLASYPVDVLOTANHLQAWISFS 403
Db 355 QVTFVFGMSDAGITNALATFFSRHGWEMAKYRVFLGSGFPHDHFESLNYLHLYTVSFS 414

QY 404 YGKKPYMSVYLHTE-----AFS 422
Db 415 YRKNKPYLSVYLHSPFETGQWPAFS 438

RESULT 4
Q9C140
ID Q9C140 PRELIMINARY; PRT; 441 AA.
AC Q9C140
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE Putative dimethyl-allyl-tryptophan-synthase (Fragment).
GN CPD1.

```

OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
OC	Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipes.	
OX	NCBI_TaxID=5111;	
RN	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RA	STRAIN=75;	
RC	Correia T., Tudzynski P.;	
RT	"Molecular analysis of dimethyl-allyl-tryptophan-synthase-genes.";	
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AJ312754; CAC37397.1; --	
FT	NON_TER 441	
SQ	SEQUENCE 441 AA; 50690 MW; ECD342B080D42FE CRC64;	
Query Match 36.6%; Score 849.5; DB 3; Length 441;		
Best Local Similarity 41.8%; Pred. No. 1.7e-63;		
Matches 172; Conservative 70; Mismatches 158; Indels 11; Gaps 4;		
QY	15 FVYLSQALNLSKDHKXWYSTAPFATMMAGAGYDVHQAQYKFLCIHREVIIPALGPYPE 74	
DB	12 YEILSLIFDPNEQRLWHSTAPFAMLDNAGYNIHDOYHGLGIFKXHIIPFLGVPT 71	
QY	75 KGQPMHWKSHLTRFGLPFELSNYSKSLRFAPEPLGSLTGKODPFNTQAIRPVLDLK 134	
DB	72 KDKE-RWLRSILTRGLPSELNCDSDSVVRYTYEPTNEVTGTEKDPFNTLAIMASVOKLA 130	
QY	135 AMVPGLDLWDFHFTKALVUSSEEARLLDRIEIPVFTONKLAADLEPSGDIVLKVYI 194	
DB	131 QIQAGIDLEWFSYFKDELTDSESATLSQSELVKEQIKTKQNKLAIDLES-QFALKVYF 189	
QY	195 YPRIKSIATGTPKRIELMFDAL-KAADKFGKVATPLAILEEFIAERAPT-----LIGH 245	
DB	190 YPHKLSIATGKSTHDLIFDSVFKLSQKSDSIQAPXQVLCYVSRNNSAESQDIALHAR 249	
QY	246 FLSCDLVKSSEHIKYCHERQDLASIEGIWTLGRNDPDTLGLDLRELWQLLPVT 305	
DB	250 LLSCLDIDPAKSVXYLLEKTVLSVMEDLWTLGGQVDASTMDGLDLRLSLLKVP 309	
QY	306 EGLCPLPNCPEYEGTSPQEQLPPIINFTLSPKSALPEPQIYFPFAFGQNDKTIAGELATPF 365	
DB	310 TGHLEVPKGLVLEGEIPNEQLPSMANYTLHNNMPPEQVYTFVFGMNDAAEISNALTIFP 369	
QY	366 ESRGWGLAKSPADIASYYPDVLQTAHNLQWISFSGKPKYNSVLH 416	
DB	370 QRHGFDMAKKYVFLQDSYPYHDFESLNYLHAYISFYWPVGESTELYQH 420	
RESULT 5		
ID	Q9C451 PRELIMINARY; PRT; 187 AA.	
AC	Q9C451;	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	Dimethylallyl tryptophan synthase (Fragment).	
GN	PAXD.	
OS	Penicillium paxilli.	
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.	
OX	NCBI_TaxID=70109;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21106005; PubMed=11169115;	
RA	Young C., McMillan L., Teifer E., Scott B.;	
RT	"Molecular cloning and genetic analysis of an indole-diterpene gene	
RT	cluster from Penicillium paxilli."	
EL	Mol. Microbiol. 39:754-764 (2001).	
DR	EMBL; AF279808; AAK11526.1; --	
FT	NON_TER 1	
SQ	SEQUENCE 187 AA; 21469 MW; 6D303E02B8977FA3 CRC64;	
Query Match 8.7%; Score 202.5; DB 3; Length 187;		
Best Local Similarity 30.5%; Pred. No. 4.2e-09;		
Matches 50; Conservative 33; Mismatches 66; Indels 15; Gaps 5;		
QY	257 SRIKYCMERQDLASIEGIWTLGRNDPDTLGLDLRELWQLLPVTIEGLPLNCFY 316	
DB	12 TRVKIYFATQSTAFNNVDIFLGGRLDGFEMORATKELKLMW-----STWAIPLGLR 65	
QY	317 EPGTSPQEQLP---FIINFTLSPKSALPEPQIYFPA--FGQNDKTIAGELATPFESRGWG 371	
DB	66 DDETLKPSPLPCAGVIFNEIPWFGADKPNPKIYPCAYTGKDDLDIADGMSDFKQGWNS 125	
QY	372 GLAKSPAD-LASYYPDVLQTAHNLQWISFSGKPKYNSVY 414	
DB	126 KSFHSYKNDYIKAFVXGDKVMCRHH---DISFSYKGGAYITAY 166	
RESULT 6		
ID	Q8S922 PRELIMINARY; PRT; 577 AA.	
AC	Q8S922;	
DT	01-JUN-2002 (TrEMBLrel. 21, Created)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)	
DE	Cell wall invertase (EC 3.2.1.26).	
GN	OSCIN1.	
OS	Oryza sativa (japonica cultivar-group).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
OC	Ehretioideae; Oryzaceae; Oryza.	
OX	NCBI_TaxID=39947;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	STRAIN=cv. Nipponbare; TISSUE=Panicle;	
RC	MEDLINE=21975747; PubMed=11978873;	
RA	Hirose T., Takano M., Terao T.;	
RT	"Cell wall invertase in developing rice caryopsis: Molecular cloning	
RT	of OSCIN1 and analysis of its expression in relation to its role in	
RT	grain filling."	
RL	Plant Cell Physiol. 43:452-459 (2002).	
CC	-!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.	
DR	EMBL; AB073749; BAB90855.1; --	
DR	Gramene; Q8S922; --	
DR	GO; GO:0004564; P:beta-fructofuranosidase activity; IEA.	
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.	
DR	InterPro; IPR001362; Glyco_hydro_32.	
DR	Pfam; PF00251; Glyco_hydro_32; 1.	
DR	SMART; SM00640; Glyco_32; 1.	
DR	PROSITE; PS00609; GLYCOSYL_HYDROL_F32; 1.	
KW	Glycoprotein; Glycosidase; Hydrolase.	
SQ	SEQUENCE 577 AA; 63984 MW; 41FE6384284F97F1 CRC64;	
Query Match 5.0%; Score 117; DB 10; Length 577;		
Best Local Similarity 24.7%; Pred. No. 0.41;		
Matches 111; Conservative 50; Mismatches 152; Indels 136; Gaps 29;		
QY	62 REVIIPALGPY--PEKGQPM-----HWKSHL-----TRFGLPFELSFNYSK 100	
DB	165 REWVKPYNVATPEPGNNATQFRDPTTAVADGHRMLVGLGKGLAYLYRSRDFK 224	
QY	101 SLLRPAPFPLGS-LTGTKDDP--FNTQAIRVLQ-DLKAWP-----GLDLWDFDH 148	
DB	225 TWVR-AKHLHSALTGMWECDFPFLQA--FGLQGLDTSVPSSKYVLKNSLDLTRYDY 281	
QY	149 TKALVUSSEEARLLDRIEIPVFTONKLAADLEPSG-----IVLKYIYYP- 196	
DB	282 T-----VGIYNKVTYRYPDPNPGDYHRLRYDYGNFYASKTFDFPV 322	
QY	197 RIKSTATGTPKRLMFDALKAADKFGKVATPLAI-----LEEFIAERAPTLGHFLS 248	
DB	323 KHRILLGWANESDSVTYDKAKGWAGIHAIPKRWLDPSGKQLQWPEELETLRKGSVS 382	
QY	249 C--DLVKPSE---SRIKY---CMERQDLASIEGIWTLGRNDPDTLGLDL- 295	
DB	383 VFDKVVKPGHEFQVIGLTGYQADVEVLSVGLKAEAL-----DPAFGDNAERLCGAKG 437	

QY 296 -----RELQQLPVTEGICPLNCF---YEGTSPQEQLPFIINFITLSPKSALEPQ 344
 DB 438 ADVRGVVGGLWL--ASAGLEXTAVFRVFKPAGHAE--PVVLMCTDPTKSSL-SPD 492
 QY 345 IYFPAP-GONDKTIAEGLATF-----PESRWGG-----LAKSYPA-----DIASY-- 385
 DB 493 LYKTFAGVTDIISGKISLRSIDRSVVSFGAGGKTCILSRVPSMAIGDKAHLYVF 552
 QY 386 --PDVDLQTAHQLQAWISFSYKGGKPYMS 412
 DB 553 NNGEADIK-ISHLKAW-----EMKKPLMN 575
 RESULT 7
 Q96X17 PRELIMINARY; PRT; 1772 AA.
 ID Q96X17, AC Q96X17,
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Sec7p.
 GN SEC7.
 OS Pichia pastoris (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4922;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Soderholm J.F., Bevis B.J., Glick B.S.;
 RT "A vector for pop-in/pop-out gene replacement in Pichia pastoris.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF356651; AAK40234.1;
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR00904; Sec7.
 DR Pfam; PF01369; Sec7; 1.
 DR SMART; SM00222; Sec7; 1.
 DR PROSITE; PS00190; Sec7; 1.
 SQ SEQUENCE 1772 AA; 200298 MW; 3BD09DF4136C8888 CRC64;

Query Match 4.8%; Score 112; DB 3; Length 1772;
 Best Local Similarity 20.4%; Pred. No. 5.8;
 Matches 95; Conservative 62; Mismatches 150; Indels 159; Gaps 23;
 QY 49 YDHAQYKFLCIHREVIIPALGPYKPGQPMHWKSHLTRGLP-FELS----FNYSKSL 103
 DB 525 YDCDSQLPSLC---EGLI-----DYLTFSTLRVEISQOQKINFRASLT 565
 QY 104 R-----FAPEPLGSLTGKDDPN-----TQAIRPVQLDKAMVPGL 140
 DB 566 RSLAVYSLKQSPMLSKLIGANVPDPEASYNPFGEYAVIESIECVLVLSLST----- 620
 QY 141 DLEWDFHTK-ALVWSEEEARTLDRDIEIPVKTQNKLAADL-RPSGDIVLKTIVYPI 198
 DB 621 ---WVDSVAKQAVESEEDTALSNGAGEDEILSORSESATOLSETG-----IPDP 670
 QY 199 KSIATGTPKERLWFDKAADKFGKVATPLAILEEFIAERAP-----TLL 243
 DB 671 AKFDTQKQRTALFSCAFNFKPKIGIAKAIKESGFIKDDSPQETAKFELLVTDGLDKTQI 730
 QY 244 GHFLSCDLVKPSESRIKVKYCMERQDLASIEGIWTLNGRNDPETHLDGLDALRELWQ--- 300
 DB 731 GEYLG-----EGDEKNITI---MDFVLDLMDFSG-----LEFVEAMRTFLQNR 771
 QY 301 -----LLPVTEGLCPLNCFYEPGTSPQEQLPFIINFITL-----SPKSALEPQ 344
 DB 772 LPGESQKIDRFMLKEARFP-----VLNNGPTFANADVFLVLAISVILLNTDQHSQAVKR 826
 QY 345 IYFPAGGONDKTIAEGLATFFESRWGGGLAKSYRADLAS--YYP-----DVDLOT----- 392
 DB 827 MSLOQFIRNAGIDG-----QDLPETLSKIYTYIQSNEIKLOEQQAAL 872
 QY 393 -ANHLO-----AWISFSYKGGKPYM-----SVYLHTFEAFSAQAQ 427

DB 873 LAGHIQPEPVTGLFSFRNQEREQYMQLSKELTLNTEKVKFSFGQE 917
 RESULT 8
 Q8TXF1 PRELIMINARY; PRT; 462 AA.
 ID Q8TXF1, AC Q8TXF1,
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE CO dehydrogenase/acyl-CoA synthase gamma subunit (Corrinoid Fe-S protein).
 DE protein.
 GN CDHE OR MK0723.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 OC Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natale D.A., Rogozin I.S., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozhavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
 DR EMBL: AE010365; AAM01937.1;
 DR GO; GO:0006730; P-one-carbon compound metabolism; IEA.
 DR InterPro; IPR004486; CdhD.
 DR InterPro; IPR007202; Fes.
 DR Pfam; PF03599; CdhD; 1.
 DR Pfam; PF04060; Fes; 1.
 KW Complete proteome.
 SQ SEQUENCE 462 AA; 50652 MW; 1E97B54407368D3B CRC64;

Query Match 4.8%; Score 110.5; DB 17; Length 462;
 Best Local Similarity 27.0%; Pred. No. 1;
 Matches 62; Conservative 28; Mismatches 91; Indels 49; Gaps 12;
 QY 125 AIRVLQDLKAMVGLDLEWDFHTKALVWSEEEARTL-LDRDIEIPVKTQNKLAADL 182
 DB 159 AVALITTDPKVMEAGLDV--FDERPLYPATEENVEDLAKLAADGCPPL-----GLHARDV 212
 QY 183 EPSGDIVLKTIVYPIKSIATGTPKERLMPDAIKAAADKFGKVATPLAILEEFIAERAPTL 242
 DB 213 EDLVPLVVEAQYQYTDLLDLPGT--BFGPHDVVSTTDKLAIEIRK--AAIEEPESFGYPTL 268
 QY 243 LG-----HFLSCDLVKPS--ESRIKVKYCMERQDLASIEGIWTLNGRNDPETHLDGLDALR 296
 DB 269 VTFEPVAFLEDDPVKAARRESYLAACVRYADILIMDV----- 308
 QY 297 ELMQLLPV-TEGLC-----FLPNCFYBPGT-----SPQEQLPFIINFITL 335
 DB 309 EPWALLPVLTRQCVVTDPREPQVEPGLVYRIGDPDENSEPVLVTTNFTLT 358

RESULT 9
 Q21549 PRELIMINARY; PRT; 490 AA.
 ID Q21549, AC Q21549,
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE M176.2 protein.
 GN M176.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.

OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
ON	[1]_RN	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TISSUE=Forelimb;	
RX	MEDLINE=22354683; PubMed=12466851;	
RA	The FANTOM Consortium,	
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;	
RT	"Analysis of the mouse transcriptome based on functional annotation of	
RT	60,770 full-length cDNAs";	
RL	Nature 420:563-573(2002).	
DR	EMBL; AK031059; BAC27231.1; -.	
DR	MGI; MGI:106040; Cyp51.	
DR	GO; GO:0004295; F-tryptsin activity; IEA.	
DR	GO; GO:0006118; Pielectron transport; IEA.	
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.	
DR	InterPro; IPR001128; Cytochrome_P450.	
DR	InterPro; IPR001254; Peptidase_S1.	
DR	Fram; PF00067; P450; 1.	
DR	PRINTS; PR00385; P450.	
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.	
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.	
SQ	SEQUENCE 503 AA; 56745 MW; 651B4F350B4DC0C7 CRC64;	

Query Match	4.3%;	Score 101;	DB 11;	Length 503;
Best Local Similarity	19.1%;	Pred. No. 7.6;		
Matches	97;	Conservative	68;	Mismatches 196;
			Indels 148;	Gaps 21;

QY	10	LLEPKPVYLSQALNSKDKTKWYSTAPFWATFWAGAGYDVAAQYKF-LCIHREVIIP-	67
DB	6	LLQSGGVLQAME-----QVTGGNLSLLLIACAFTLISLVYLFRLAVGHMWQLPA	56
QY	68	-ALGPYPEKQPMHWKSHLTRFG-LPELSFN-YKSLRLFAPEPLGS-----	112
DB	57	GAKSP-PHYSPYFFLGHATAFGKSPFIEFLNAYEKYGPVFSPTMVGKTFYLLGSDAAA	115
QY	113	-LTGTKDDPNTQAI-----RPV-----LQDLKAMVPGLDLEWDFDHTKAL	152
DB	116	LFENSKNEDLNAEVYGRLLTPFGKVAYDVPNAIFLEQKKIKSLNIA---HFQYV	172
QY	153	VVSEEAR-----TLDRDIEIPVFKTONKLAADLEP	184
DB	173	PIIEKAKEYFQSGWSEGRNVFEALSELIIITASHCLHGKEIRSQLNERKVAQIYADLD-	231
QY	185	SGDIVLKTYYP-----RIKSI-----ATGTPKRLMFDKAAAD	219
DB	232	-GGFAHAAMLLPAWLPLSPFRDRRAHREIKNIIFYKAIQKRRLSKEPAEDILQTLDSY	290
QY	220	KFKGVATP-----LAILEEFIAERAPTLGLHFSLCDLVKPSERI----KVYQWE-	265
DB	291	KDGRPLTDEEISGMVLIGLLAGQSTSTTSAAWGGFFLAKD--KFLQEKCYLEQKAVCGED	348
QY	266	-----RQDLASIE-----GIWTLNGRNDPETHDGLDALRELWQLLPVTEG	307
DB	349	LPPLTYDQLKDLNLLORCIKETURLRPPIIMTMRMAKTPQTAVG-----YTIPEGHQ	400
QY	308	LCPLPNCFYBFGTSPQQLPFIINFTLSPKSALPEQIYFPAPFQNDKTTIAEGLATPFES	367
DB	401	VCVSPVTNQRKDSWAERLDFNDRYLQDNFASGEKFAYYVFFGAGRHCRCVGENFAYVQIK	460
QY	368	RWGGLAKSYPADLAS-YYPDDVLQTAHH	395
DB	461	TIWSTMRLRYEFDLNGYFFFTVNTTMIH	489

RESULT 14				
Q42710				
ID	Q42710	PRELIMINARY;	PRT;	878 AA.
AC	Q42710;			
DT	01-NOV-1996	(TREMBlrel. 01, Created)		
DT	01-NOV-1996	(TREMBlrel. 01, Last sequence update)		

DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)	
DR	Lipoxigenase [EC 1.13.11.12].	
CS	Cucumis sativus (Cucumber).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;	
OC	eurosid1; Cucurbitales; Cucurbitaceae; Cucumis.	
CC	NCBI_TaxID=3659;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Cotyledon;	
RA	Medline=97054894; PubMed=889881;	
RX	Hoehne M., Nellen A., Schwennesen A., Kindl H.;	
RT	"Lipid body lipoxigenase characterized by protein fragmentation, cDNA	
RT	sequence and by its very early expression during germination of	
RT	cucumber seeds.";	
RL	Eur. J. Biochem. 241:6-11 (1996).	
DR	EMBL; X92890; CAA63483.1; -.	
DR	PIR; S74207; S74207.	
DR	HSP; P08170; 2SBL.	
DR	GO; GO:000506; F:iron ion binding; IEA.	
DR	GO; GO:0016165; F:lipoxigenase activity; IEA.	
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.	
DR	GO; GO:0006118; P:electron transport; IEA.	
DR	InterPro; IPR001097; Lipoxigenase.	
DR	InterPro; IPR001024; Lipoxigenase_LH2.	
DR	InterPro; IPR008976; PLAT LH2.	
DR	Pfam; PF00305; lipoxigenase; 1.	
DR	Pfam; PF01477; PLAT; 1.	
DR	PRINTS; PR00087; LIPOXYGENASE.	
DR	SMART; SM00308; LH2; 1.	
DR	PROSITE; PS00711; LIPOXYGENASE 1; 1.	
DR	PROSITE; PS00081; LIPOXYGENASE_2; 1.	
DR	PROSITE; PS00095; PLAT; 1.	
DR	Oxidoreductase.	
KN	SEQUENCE 878 AA; 99768 MW; A92660AF4D1ED3BE CRC64;	
SQ		
	Query Match 4.3%; Score 101; DB 10; Length 878;	
	Best Local Similarity 21.7%; Pred. No.17;	
	Matches 94; Conservative 49; Mismatches 129; Indels 162; Gaps 26;	
QY	61 HREVIIPALG-----PYPEK---GQPMHWKSHLTREGLPFELSFNYKSLIRFAPEPLCS 112	
DB	235 HR-----PILGGTTEYPYPRGTRGPRSRDH-----NYESRL-----SPINS 273	
QY	113 LT--GPKDDPFN-----TQAIRFVLQDKAMVPLDLEWDFHFTKALVVS 155	
DB	274 LDIVYKDFNGFHLKMSDFLYTLKALSIKFGLSIFDVT-----NEFDNF----- 322	
QY	156 EBEARTLLDRDIEP--VPKTONKAADLEPSGDIVLKYTVPRKSIATGTPKRLMFD 213	
DB	323 -KEVDNLFRGFPFPNFAKT---ITEDLTP-----PLFKALVRNDGKFLKFP 367	
QY	214 AIKAADKFGKV--ATPLALEBFIABRAPTLGHFLSCDLVPKSESRIKVCWGEROLDLA 271	
DB	368 TPENV-KDNKGNSTDEEFAREMLAGPNLLI-----RRLEAFPTSKLDP- 412	
QY	272 STEGIWTLGRNDPPT-----LDGLDALRELWQ-----LLPVTTEGLCPPLNCFVEPQT 320	
DB	413 -----NVYGNQNSTITESHIKHGLDGLTVDEAMQNRLYIVDFHDALPYLTRMNATST 466	
QY	321 SPOEQLPFI-----NFTLSP---KSALPERQ-----IYPAFGQNDKTTAEGLATFF 365	
DB	467 KTVATETLLLLKDDGTLKPLVIELALPHFPQGGOLGAISKLYPFAENGVOKSI----- 518	
QY	366 ESRGWGLAKSPADLASIYPPVDQLTANHLQAWISFSYKPKPKYMSVYLHT---FEAFS 422	
DB	519 ----W-QLAKEY-----VTVDVGVHQLSHLWHTHAVLEPFPV 551	
QY	423 AAA-QEVAMCHDH 435	
DB	552 IATHROLSVLPHIH 565	

